THE INSTITUTE OF PHYSICS
AND
THE PHYSICAL SOCIETY

REPORTS ON PROGRESS IN PHYSICS

VOLUME XXIV (1961)

EXECUTIVE EDITOR
A. C. STICKLAND

Published by
THE INSTITUTE OF PHYSICS AND THE PHYSICAL SOCIETY
47 Belgrave Square, London, S.W.1
Editorial Office: 1 Lowther Gardens, Prince Consort Road, London, S.W.7
1961

Printed in Great Britain by John Wright and Sons Ltd.
The Stonebridge Press, Bath Road, Bristol 4

CONTENTS

	PAGE
Preface	iv
The Theory of the Superconductive State. By H. Fröhlich	1
High Current Gas Discharges. By A. A. Ware	24
Cosmic Radio Waves and their Interpretation. By J. L. Pawsey and E. R. Hill	69
Magnetic Domains. By D. J. Craik and R. S. Tebble	116
Photoelectronic Image Intensifiers. By J. D. McGee	167
Ferrimagnetism. By W. P. Wolf	212
Theory and Applications of the Density Matrix. By D. ter Haar	304
The Dynamics of High Temperature Plasmas. By W. B. Thompson	363

PREFACE

The *Reports on Progress in Physics* is under the general supervision of a special Sub-committee of the Publications Committee of the Institute of Physics and the Physical Society. The Members of the Sub-committee during the 1960–61 session were:

Chairman of the Publications Committee: Professor R. W. Ditchburn.

Chairman of the Sub-committee: Dr. M. Blackman.

Members of the Sub-committee: Dr. K. W. Allen, Dr. B. R. Coles, Professor B. H. Flowers, Dr. R. Latham, Dr. P. T. Matthews, Dr. K. Mendelssohn, Dr. J. M. Ziman.

Reports published in the Volume are also being issued as individual articles. Information may be obtained from the offices of the Institute and Society.

A. C. STICKLAND,
Editor and Deputy Secretary.

Vol. 24
1961

THE THEORY OF THE SUPERCONDUCTIVE STATE

By H. FRÖHLICH
Department of Theoretical Physics, University of Liverpool

CONTENTS

	PAGE
§ 1. Preface	1
§ 2. Introduction—Phenomenological descriptions	2
2.1. The London equations	3
2.2. The gap	5
§ 3. The introduction of field theory into solid state physics	6
3.1. Bloch's theory of conductivity	8
3.2. Field theoretical aspect	9
§ 4. Isotope effect and gap	10
4.1. Classification of interaction	10
4.2. The isotope effect	11
4.3. The gap, one-dimensional	12
4.4. Pairs and the gap	14
§ 5. Criticisms; Meissner effect	15
5.1. Gauge invariance	16
5.2. Other criticisms	18
5.3. Meissner effect	19
§ 6. The psychology of superconductivity	20
§ 7. Outlook	21
References	22

Abstract. The development of the theory of the superconductive state is described in terms of simple physical ideas so far as this has been found possible.

§ 1. Preface

Amongst the theories on properties of materials the theory of superconductivity has played a unique role. Discovered about fifty years ago, this phenomenon defied theoretical interpretation for about forty years even when it appeared that the basic equipment should be available. The present article does not intend to give a detailed account of the present state of the theory. Good articles on this are available elsewhere (Kuper 1959, Schafroth 1960, Bardeen and Schrieffer 1961, Bardeen 1956 for earlier stages); others will no doubt appear in the near future. Instead, I intend to recapture some of the excitement and fascination that surrounded the development of the theory, and to explain what has been achieved. This aim implies that a physical picture of the various steps should be given without the use of advanced mathematical language. Unfortunately, this was not possible in the most recent steps, and I feel that this means that these steps are not at present clearly understood. It was necessary, therefore, to outline in § 5 these recent developments in a slightly more mathematical language, so as to be able to express some criticisms which I hope may ultimately help to arrive at a simple physical picture.

This article will be of interest, I hope, to the non-specialist. It addresses itself, however, also to specialists, in particular to those who excel in the use of a single technique such that their understanding has the form of a δ-function. There is hope that some of them might realize that a calculation should be preceded by an idea.

With regard to references, I have quoted all papers in which an important step is published for the first time, but not necessarily subsequent improvements. While such improvements naturally give a more complete insight into a particular situation, I think that first steps normally deserve considerably more credit than developments arising from them with the use of established methods. First steps usually are the action of a thinking being, while developments arising out of them in principle might be carried out by some superb calculating machine.

§2. Introduction—Phenomenological Descriptions

Superconductivity was discovered in 1911 by Kamerlingh-Onnes (1911), who was the first to liquefy helium and thus to reach lower temperatures (below 10°K) than had been obtained hitherto. He investigated the temperature dependence of the electric resistivity of metals and found that abruptly at a temperature T_c the resistance of some of them decreased to zero and remained zero below T_c. This discovery was, no doubt, startling. In trying to imagine the impact it made on physicists it must be remembered, however, that the properties of the electric conductivity of metals were not well understood then, even at high temperatures. Also, in those days, many other discoveries were made that could not be understood on the basis of existing theories. Superconductivity thus did not occupy a separate place.

Kamerlingh-Onnes in pursuing his investigations found already in 1913 that superconductivity is destroyed by sufficiently strong magnetic fields. As in the temperature dependence, a sharp transition was obtained at a critical field H_c. It was not until 1933, however, that Meissner and Ochsenfeld (1933) investigated the magnetic properties of superconductors in some detail. They discovered that the magnetic induction in a superconductor vanishes (Meissner effect)—it represents a perfectly diamagnetic material. It must be remembered now that in 1933 the situation with regard to the understanding of electric conductivity, and of other properties of materials, was radically different from that in 1911. Based on the work of Sommerfeld (1928) and Bloch (1928) nearly all properties of materials, and in particular of metals, were understood at least in a qualitative, if not in a quantitative, way. The temperature dependence of the resistivity, and the magnetic properties of normal metals in particular, were very well understood. Yet the theory appeared to be completely helpless in the case of superconductivity. Since it was felt in 1933 that the basic concepts of the structure of materials were sufficiently far developed to account for all of their properties, this apparent helplessness of the theory led to a general feeling of exasperation.

In this situation it was of great importance that satisfactory phenomenological theories of the thermal, as well as of the electrodynamic, behaviour of superconductors were developed. Such theories do not attempt an understanding from an atomistic point of view. They do, however, clarify the situation by showing what properties an atomistic (microscopic) theory ought to derive. The

most important step in the thermodynamics of superconductors was the realization that the superconductive state represents a different thermodynamic phase from the normal state of a metal. It followed immediately that a microscopic theory must, in the first place, explain the existence of a second phase; only then can one hope to show that this second phase has the required electrodynamic properties. Applications of thermodynamics were already made at an early stage by Keesom (1924). The discovery of the Meissner effect was, however, required to establish the superconductive state clearly as a thermodynamic phase, and an appropriate thermodynamic theory was developed by Gorter and Casimir in 1934 (also Gorter 1933). Such a theory shows that the phase transition is of the second kind (in Ehrenfest's (1933) classification), which means that at the transition temperature there is a jump in the specific heat, though no latent heat is involved in the transition. This type of phase change is characteristic for order–disorder transitions (e.g. ferromagnetism). Application of a magnetic field above the critical one H_c (which decreases with increasing temperature T, leading to $H_c \to 0$ as $T \to T_c$) permits a determination of the energy difference between the normal and the superconductive states. This difference is exceedingly small, of the order 10^{-7} ev per atom. An atomistic theory must thus be able to explain how a phase change involving such a tiny energy can result in a vast change of the electromagnetic properties.

2.1. *The London Equations*

The phenomenological theory of the electromagnetic properties of superconductors was developed by F. and H. London (1935, also F. London 1950). All materials satisfy, of course, the Maxwell equations which are expressed in terms of electric field and displacement, $\mathbf{E}$ and $\mathbf{D}$, the magnetic field and induction, $\mathbf{H}$ and $\mathbf{B}$, the conduction current density $\mathbf{j}$ and the free charge density. Properties of materials are described in terms of further connections between these quantities; $\mathbf{j} = \sigma\mathbf{E}$, for instance, represents Ohm's law in terms of the specific conductivity σ of a conductor. This law clearly must be invalid for a superconductor. F. and H. London assume that in a superconductor the current density can be split into a normal $\mathbf{j}_n$ and a superconductive $\mathbf{j}_s$ contribution,

$$\mathbf{j} = \mathbf{j}_n + \mathbf{j}_s. \qquad \ldots\ldots(2.1)$$

Here $\mathbf{j}_n$ satisfies Ohm's law but $\mathbf{j}_s$ follows the two equations

$$\Lambda \frac{\partial}{\partial t}\mathbf{j}_s = \mathbf{E}, \qquad \ldots\ldots(2.2)$$

$$c\Lambda \operatorname{curl} \mathbf{j}_s = -\mathbf{B}, \qquad \ldots\ldots(2.3)$$

where Λ is a new material constant. The first of these equations (2.2) can be interpreted as expressing the frictionless acceleration of a charge density $Q = en$ connected with a mass density $M_t = mn$ provided

$$\Lambda = \frac{M_t}{Q^2} = \frac{m}{ne^2}. \qquad \ldots\ldots(2.4)$$

Naturally, one will try to interpret e, m, n as charge, mass and number density of the conduction electrons. This first equation (first postulated by Becker, Sauter

and Heller 1933) clearly permits the existence of a current in the absence of an electric field.

The second London equation (2.3) refers to that part of $\mathbf{j}_s$ only whose curl does not vanish—we might call it the transverse part. Assuming, for simplicity, $\mathbf{B} = \mathbf{H}$ and stationariness, Eqn (2.3) together with the Maxwell equations

$$\operatorname{curl} \mathbf{H} = 4\pi \mathbf{j}/c, \quad \operatorname{curl} \mathbf{E} = 0 \text{ in the stationary case} \qquad \ldots\ldots(2.5)$$

yields

$$\nabla^2 \mathbf{H} = \frac{4\pi}{c^2 \Lambda} \mathbf{H} \qquad \ldots\ldots(2.6)$$

for the magnetic field. Solutions for particular shapes of a superconductive material (cf. London 1950) show that an applied magnetic field penetrates the surface only to a depth of the order $c(\Lambda/4\pi)^{1/2}$ which for typical values of n is of the order of 10^{-5}–10^{-6} cm. Extensive experiments (cf. Shoenberg 1952) have confirmed these predictions of the London theory—though it was found that details require certain modifications (Pippard 1953).

The second London equation (2.3) can be expressed in an alternative manner by introducing the vector potential $\mathbf{A}$ from which the magnetic induction $\mathbf{B}$ can be derived by

$$\mathbf{B} = \operatorname{curl} \mathbf{A}. \qquad \ldots\ldots(2.7)$$

From (2.7) and (2.3) it follows that

$$c\Lambda \mathbf{j}_s = -\mathbf{A} \qquad \ldots\ldots(2.8)$$

provided the gauge of the vector potential is chosen such that

$$\operatorname{div} \mathbf{A} = 0, \quad \text{and} \quad A_\perp = 0, \qquad \ldots\ldots(2.9)$$

where $A_\perp$ represents the boundary value of $\mathbf{A}$ normal to the surface of the superconductor.

In the form (2.8) the second London equation can be interpreted in terms of microscopic properties. Let $\mathbf{p}$ be the momentum of an electron. In the presence of a vector potential $\mathbf{A}(\mathbf{x})$ ($\mathbf{x}$ = space coordinates) its velocity $\mathbf{v}$ is given as

$$m\mathbf{v} = \mathbf{p} - \frac{e}{c}\mathbf{A}. \qquad \ldots\ldots(2.10)$$

Quantum field theory describes the properties of an assembly of electrons in terms of a field $\psi(\mathbf{x})$ and its conjugate $\psi^*(\mathbf{x})$. Details of the operational properties of $\psi(\mathbf{x})$ are irrelevant at present; we notice, however, that

$$n(\mathbf{x}) = \psi^*(\mathbf{x})\psi(\mathbf{x}) \qquad \ldots\ldots(2.11)$$

represents the operator for the number density and, following (2.10),

$$\mathbf{j}_{op}(\mathbf{x}) = \frac{e}{m}\psi^*(\mathbf{x})\left[\mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{x})\right]\psi(\mathbf{x}) \qquad \ldots\ldots(2.12)$$

represents the operator for the current density. Forming now expectation values and assuming $\mathbf{A}(\mathbf{x})$ to be slowly varying, then

$$\mathbf{j} = \langle \mathbf{j}_{op} \rangle = -\frac{e^2}{mc}\langle \psi^*(\mathbf{x})\mathbf{A}(\mathbf{x})\psi(\mathbf{x})\rangle = -\frac{e^2 n(\mathbf{x})}{mc}\mathbf{A}(\mathbf{x}), \qquad \ldots\ldots(2.13)$$

provided it is assumed that

$$\langle \psi^*(\mathbf{x})\mathbf{p}\psi(\mathbf{x})\rangle = 0. \qquad \ldots\ldots(2.14)$$

Equation (2.13) is identical with (2.8) using Λ from (2.4). A microscopic interpretation of the second London equation thus requires that (2.14) holds provided the vector potential is gauged according to (2.9). Now assume conditions in which $\mathbf{A}(\mathbf{x})$ can be treated as a small perturbation. (Eqn (2.9) ensures this for sufficiently weak fields.) In an equilibrium state in the absence of a field ($\mathbf{A} = 0$) (2.14) clearly holds. Application of a field in a non-superconductor then leads to a perturbation of this state such that the left-hand side of (2.14) is different from zero. Condition (2.14) for superconductivity, on the other hand, requires that—in first order in $\mathbf{A}$—the state of the system is not perturbed by the field. A similar condition is satisfied by a single diamagnetic molecule placed in a magnetic field—but not for a system of such molecules. A superconductor thus behaves like a diamagnetic molecule of macroscopic dimensions.

2.2. *The Gap*

The condition that the state of the system should not be changed in first order in the vector potential $\mathbf{A}$ (gauged by (2.9)) can also be expressed in terms of the change of energy due to the application of the magnetic field. We note that both with and without a magnetic field the energy of an electron can be written as $\frac{1}{2}mv^2 +$ potential energy, if $m\mathbf{v}$ is given by (2.10). Introducing the electron field $\psi(\mathbf{x})$, similar to the use in (2.12), the total energy contains two terms in $\mathbf{A}$, one of first order (symmetrize $\mathbf{pA}$)

$$H_A = -\frac{e}{mc}\int \psi^*(\mathbf{x})\,\mathbf{pA}(\mathbf{x})\,\psi(\mathbf{x})\,d^3x \qquad \ldots\ldots(2.15)$$

and one of second order

$$H_{AA} = \frac{e^2}{2mc^2}\int A^2(\mathbf{x})\,\psi^*(\mathbf{x})\,\psi(\mathbf{x})\,d^3x = \frac{e^2}{2mc^2}\int A^2(\mathbf{x})\,n(\mathbf{x})\,d^3x. \qquad \ldots\ldots(2.16)$$

For non-superconductive materials the expectation values of H_A and of H_{AA} nearly cancel (they cancel exactly in classical physics). For superconductors, however, the London condition (2.14) implies that

$$\langle H_A\rangle = 0. \qquad \ldots\ldots(2.17)$$

To calculate this expression in a formal way by perturbation theory, we assume the system to be in the ground state (energy E_0) with the perturbation H_A inducing virtual transitions into states with energy E_n connected with matrix elements M_{0n}. The resulting energy perturbation due to H_A is then

$$\langle H_A\rangle = \sum_n \frac{|M_{0n}|^2}{E_0 - E_n}. \qquad \ldots\ldots(2.18)$$

In a normal metal the matrix elements M_{0n} refer principally to the virtual excitation of single electrons by the energy $E_n - E_0$. This energy difference can be arbitrarily small so that the main contribution to (2.18) arises from the very small denominators $E_n - E_0$. Assume now that in a superconductor (i) a finite minimum energy Δ is required for single electron excitations (an energy gap Δ exists) then the magnitude of (2.18) will be reduced very substantially provided

(ii) the M_{0n} remain unaltered; the London condition (2.14) and hence superconductivity would thus follow from conditions (i) and (ii). This gap model was first proposed by H. Welker (1938). Unfortunately, the two conditions are contradictory, as was shown by M. J. Buckingham (1957). A simple derivation of this contradiction follows from the uncertainty relation

$$(pq-qp)_{rs} = -i\hbar\delta_{rs} \qquad \ldots\ldots(2.19)$$

or

$$\sum_n (p_{rn}q_{ns}-q_{rn}p_{ns}) = -i\hbar\delta_{rs} \qquad \ldots\ldots(2.20)$$

where p_{rn} and q_{rn} are the matrix elements of a component of momentum and displacement respectively. For nearly all models the operator equation $p = m\dot{q}$ holds which means that $\hbar p_{rn} = -im(E_r-E_n)q_{rn}$ or with (2.20),

$$m\sum_n q_{rn}q_{ns}(E_r+E_s-2E_n) = \hbar^2\delta_{rs}. \qquad \ldots\ldots(2.21)$$

Thus any change in the energy spectrum requires an appropriate change in the matrix elements q_{rn} (and hence also of p_{rn}) so that (2.21) remains fulfilled. Creation of a gap in the energy spectrum E_n is thus in contradiction to the above condition (ii).

In spite of this drawback the gap model has been extremely useful; strong experimental evidence for the existence of the gap has been accumulated lately, and the microscopic theory has at last been able to derive it (§§ 4 and 5). The gap model assumes a gap of width Δ of the order kT_c at low temperatures, $T \ll T_c$. At higher temperatures the gap shrinks gradually tending to zero as T approaches T_c. Existence of the gap is thus the distinguishing mark between the superconductive and the normal phase. Though the setting up of a gap may require very little energy, it clearly may lead to a strong change in various electronic properties. Thus, while without a gap the electronic specific heat is proportional to T, the gap alters this into an $\exp(-\Delta/2kT)$ law, if $T \ll T_c$. This law has been established experimentally (Corak *et al.* 1956). Other experimental results like those on heat conductivity (Goodman 1953) support the gap hypothesis. The clearest and most direct experimental evidence arises from optical absorption measurements (cf. Glover and Tinkham 1956, 1958) which show clearly that electronic transitions with energy difference less than Δ are present in the normal but absent in the superconductive state.

The Meissner effect, on the other hand, is not a simple consequence of the existence of a gap (an insulator, too, has a gap) but requires more subtle knowledge of the behaviour of the matrix elements describing transitions across the gap. This must obviously be expected, for otherwise there could be no distinction between a superconductor and an insulator.

§ 3. The Introduction of Field Theory into Solid State Physics

By 1933, the year in which the Meissner effect was discovered, the electron theory of metals was already well established. It was based on the assumption that the Coulomb interaction between electrons could be replaced by an

appropriate contribution to the periodic potential (lattice period) of the ions. Theoretical justification of this 'free electron' model was provided only in 1953 by Bohm and Pines (1953). The model was, however, so successful in dealing with all properties of metals, except superconductivity, that little doubt as to the correctness of its basic assumptions could exist.

Many futile attempts to deal with superconductivity were advanced. The general opinion, prior to 1950, was that a further interaction not described by the free electron model must exist between the electrons so as to establish the superconductive state. Since the Coulomb interaction between electrons had not been dealt with properly, some hope was held out in this direction. A dictum was accepted nearly universally, namely that the ions of the lattice, in view of their large mass (compared with electrons), could play no important role in the establishment of the superconductive state. It was in the face of this opinion that the author conceived the idea (Fröhlich 1950) that just the opposite of the 'dictum' contains the truth. The establishment of this approach was closely connected with the introduction of new methods, those of field theory, into solid state physics. Use of such methods is, of course, accepted now as trivial. It is just this kind of step, however, joining different aspects of physics, which often is not seen for long times.

The first explicit application of field theory in solid state physics took place in 1950 (Fröhlich, Pelzer and Zienau 1950) and referred to a problem much simpler than superconductivity. It dealt with a single electron in a polar crystal. If all ions of the crystal are assumed rigidly fixed to their equilibrium positions, then a single electron placed into the crystal is under the influence of a periodic potential (lattice period). The behaviour of an electron in a periodic potential had been treated by F. Bloch in 1928, who showed that its energy spectrum consists of bands and that within such a band an electron behaves much like a free electron provided its mass is replaced by a certain effective mass. If now the ions of the lattice are no longer fixed rigidly to their equilibrium positions, then a further interaction of the free electron with the ions may occur, depending on the displacement of the ions (and on their polarization). This displacement of the ions represents a field, the polarization field. The waves (quanta) of this field are polarization waves ; electrons are sources of the field. A single electron has a self-energy in such a field, and in contrast to the case of the electromagnetic field it was found that this self-energy does not diverge even when it is disregarded that a shortest polarization wavelength exists. The self-energy is due to the displacement of the ions by the electron.

Many years before, Landau (1933) had assumed that the heavy mass of the ions leads then to a self-trapping of the electron. The field theoretical treatment showed, however, that this need not occur at all. In fact, the kinetic energy of the ions (the field) attached to a moving electron may be very much smaller than the kinetic energy of the electron. Clearly such a result was of great interest in connection with the 'dictum' that ions, in view of their heavy mass, should be of no importance for superconductivity and the application of appropriate field theoretical methods to metals was called for. Before describing the underlying ideas of these developments we must outline Bloch's theory of conductivity which was developed in 1928.

3.1. *Bloch's Theory of Conductivity*

As mentioned above, Bloch in 1928 showed that the energy spectrum of individual electrons in a strictly periodic potential consists of bands. An electron in a permitted energy state behaves similarly to a free one, i.e. it may carry a certain current independent of time (no scattering). According to Pauli's exclusion principle each completely defined state can be occupied by a single electron only. On the free electron hypothesis, then, in the lowest state of the system all single electron levels up to a given one are occupied. In a full band the total current vanishes ; in a conductor, therefore, there exists a band (conduction band) which is not completely filled. The number of electrons within a conduction band is usually of the same order as the number of atoms in the lattice. In a reasonably good approximation, therefore, these electrons can be treated as completely free. The Pauli principle then shows that in their lowest state these free electrons occupy all energy levels between zero and an energy ζ (the Fermi level) which usually is of the order of several electron volts.

When ions of the lattice are displaced from their equilibrium positions then the potential acting on the electrons is no longer periodic. This deviation leads to a scattering of the electrons. The thermal vibrations of the ions are connected with ionic displacements which increase with temperature ; this leads to an appropriately increased scattering of the electrons and hence to an electric resistivity which increases with temperature.

For a more quantitative treatment the lattice vibrations are considered in terms of their quantized normal modes (phonons). If $\mathbf{w}$ is the wave vector of such a mode, then $\hbar w s$ is its energy (s = velocity of sound). Bloch has shown that the interaction between these vibrations and the electrons leads to matrix elements which are linear in the lattice displacements and which conserve the total wave vector. They are connected with absorption or emission of a phonon by an electron which makes a transition from a state with wave vector $\mathbf{k}$ into a state

$$\mathbf{k}' = \mathbf{k} \pm \mathbf{w}. \qquad \ldots\ldots(3.1)$$

An actual absorption or emission process also requires conservation of energy, i.e.

$$\epsilon(\mathbf{k}) \pm \hbar w s = \epsilon(\mathbf{k} \pm \mathbf{w}) \qquad \ldots\ldots(3.2)$$

where $\epsilon(\mathbf{k})$ is the electronic energy

$$\epsilon(\mathbf{k}) = \hbar^2 k^2/2m, \qquad \ldots\ldots(3.3)$$

m being its effective mass. It should be noticed that $\epsilon(\mathbf{k})$ is of the order ζ, i.e. several electron volts, whereas $\hbar w s$ is always less than the energy corresponding to the Debye temperature, i.e. less than 10^{-1} ev. An absorption or emission process is proportional to the square of the matrix element, i.e. to the square of the ionic displacement, and hence to the thermal lattice energy. At high temperatures the ensuing scattering of electrons is nearly isotropic so that the resistivity itself is proportional to the thermal lattice energy, i.e. proportional to the absolute temperature T as found experimentally. At low temperatures (scattering no longer isotropic) a T^5 law is obtained which also is in reasonable agreement with the experimental evidence.

3.2. *Field Theoretical Aspect*

Absorption and emission of quanta by particles is one aspect of field theories. It is necessarily connected with a self-energy and an interaction energy of these particles. Bloch's successful theory of metallic conductivity thus considers one aspect of a field theory. It must necessarily lead to a self-energy and an interaction energy between the electrons. This remark was the basis of the introduction of electron–phonon interaction into the treatment of superconductivity (Fröhlich 1950). It showed that it was not necessary to find 'new' interactions between electrons by going beyond the successful free electron model; this model itself contained an interaction which so far had not yet been explored.

To pose the problem in its simplest form, we require solutions for a system of electrons interacting with a vibrational field according to the Hamiltonian (Fröhlich 1952)

$$H = H_{\mathrm{el}} + H_{\mathrm{f}} + H_{\mathrm{int}} \qquad \ldots\ldots(3.4)$$

where H_{el} and H_{f} refer to the Hamiltonians of free electrons and free phonons, and H_{int} to their interaction. All parameters entering (3.4) are relevant in properties different from superconductivity; the interaction constant, in particular, entering H_{int}, determines the magnitude of high-temperature conductivity.

More explicitly, if $\psi(\mathbf{x})$ is the wave function of the electronic field, used already in (2.12)–(2.15), then in the free electron model

$$H_{\mathrm{el}} = -\frac{\hbar^2}{2m}\int \psi^*(\mathbf{x})\,\nabla^2\psi(\mathbf{x})\,d^3x \qquad \ldots\ldots(3.5)$$

where (cf. (2.11)) $\qquad n(\mathbf{x}) = \psi^*(\mathbf{x})\,\psi(\mathbf{x}), \quad N_{\mathrm{el}} = \int n(\mathbf{x})\,d^3x, \qquad \ldots\ldots(3.6)$

if $n(\mathbf{x})$ is the number density, and N_{el} the total number of free electrons. Integrations go over the whole metal. Similarly, if $\mathbf{P}(\mathbf{x})$ is the longitudinal displacement of the lattice near a point $\mathbf{x}$, $\dot{\mathbf{P}}$ its rate of change and if M is the ionic mass, and n_i their number density, then

$$H_{\mathrm{f}} = \tfrac{1}{2}Mn_i\int [\dot{\mathbf{P}}^2(\mathbf{x}) + s'^2(\mathrm{div}\,\mathbf{P}(\mathbf{x}))^2]\,d^3x. \qquad \ldots\ldots(3.7)$$

Here s' is a velocity connected (but not identical) with the sound velocity (cf. § 4). The interaction corresponding to Bloch's matrix elements is of the simplest possible form,

$$H_{\mathrm{int}} = C\int n(\mathbf{x})\,\mathrm{div}\,\mathbf{P}(\mathbf{x})\,d^3x, \qquad \ldots\ldots(3.8)$$

where the interaction constant $C > 0$ is of the order of several electron volts.

Bloch's theory is obtained from Eqn (3.4), when the interaction H_{int} is treated as a perturbation for the purpose of calculating transitions between electronic states connected with absorption or emission of phonons. In second order H_{int} yields an energy perturbation; it was by this method (Fröhlich 1950) or by similar methods (Bardeen 1950, 1951) that the first attempts at a solution were performed. Application of perturbation theory, however, is very dangerous and may lead to wrong results. The main reason for this is the high degeneracy of the unperturbed system which practically has a continuous energy spectrum only (cf. Fröhlich 1953).

Even after this difficulty was clearly realized it took until 1957 (Bardeen, Cooper and Schrieffer 1957) to obtain a satisfactory energy spectrum, including the gap, from Eqn (3.4). For a subject which received so much attention this may be considered a long delay, in particular since (i) the method used for the solution had been developed in connection with different problems already in 1947 by Bogoliubov, and (ii) the qualitative ideas were essentially conceived already in connection with the first introduction of electron–phonon interaction in 1950.

While thus a satisfactory solution was long delayed, a most important feature of the interaction H_{int} was already discovered at the first attempt. It led to the prediction of the isotope effect (Fröhlich 1950) which was discovered soon afterwards (Maxwell 1950, Reynolds *et al.* 1950, Allen *et al.* 1950a, b). This effect shows that the energy difference between the normal and the superconductive states is for isotopes of the same element inversely proportional to the isotopic mass. It proves conclusively that the interaction H_{int} is decisive for superconductivity. The isotope effect can be understood in a simple way, as will be shown at the beginning of the next section.

§ 4. Isotope Effect and Gap

4.1. *Classification of Interaction*

The interaction term H_{int} (3.8) is proportional to the product of the electronic density $n(\mathbf{x})$ and the compressional density $n_i \operatorname{div} \mathbf{P}(\mathbf{x})$ connected with lattice displacements. This term, therefore, tends to make the fluctuations of both these densities large (opposite sign) but the other terms in the Hamiltonian counteract this tendency. Both densities are dynamic quantities. On a first consideration one will be tempted to assume the electronic density $n(\mathbf{x})$ to follow adiabatically the displacement density $n_i \operatorname{div} \mathbf{P}(\mathbf{x})$ because of the low electronic mass. This, in fact, is the main effect of the interaction terms, though from the point of view of superconductivity it is not very interesting. Its main result is to modify the density vibrations of the ions which now carry a certain amount of electronic density with them. This reduces the sound velocity to a value s less than s' (s is the experimental value).

We have just argued that the electrons, in view of their low mass, follow the lattice displacements adiabatically. However, as already indicated in § 3, in connection with self-trapping, such an argument may be misleading. To show this, consider an electronic density wave obtained by superposition of two electronic states with wave vectors $\mathbf{k}$ and $\mathbf{k}'$ and energies (cf. (3.3)) $\epsilon(k)$ and $\epsilon(k')$. This density wave has a frequency $\omega = [\epsilon(k') - \epsilon(k)]/\hbar$ and a wave vector $\pm \mathbf{w} = \mathbf{k}' - \mathbf{k}$. It only interacts with lattice vibrations with the same wave vector $\mathbf{w}$, for otherwise the integral (3.8) vanishes. In general, the phonon frequency differs from ω. Now, as in the case of coupled oscillators, the electronic density follows nearly adiabatically the lattice vibrations provided ω is sufficiently larger than ws, i.e. provided $\epsilon(k') - \epsilon(k)$ is larger than $\hbar ws$, the energy of a phonon. Very many combinations of electronic states satisfy this requirement. Another case does exist, however, when

$$|\epsilon(k') - \epsilon(k)| < \hbar ws. \qquad \ldots\ldots(4.1)$$

In this case the opposite holds: the lattice displacements follow the electronic density adiabatically.

Vol. 24
1961

It is this latter case which is of interest to us. The original idea that the superconductive state should be connected with electron–phonon interaction was conceived in terms of it in the following way: An electron density causes a certain lattice displacement and thereby leads to a decrease in energy. Other electrons will find it favourable to adapt themselves to this displacement and in turn lead to a further increase in it. A situation might thus arise which could be compared with that in other second-order phase transitions like, for instance, ferromagnetism where a few oriented spins tend to orient others into the same direction which in turn increases the tendency for orientation.

The separation of the two types of interaction is already evident in perturbation theory. A more elegant manner of performing this is to redefine the lattice displacement $\mathbf{P}$ such that it contains electronic contributions following it nearly adiabatically, and to redefine the electron field ψ such that it contains the lattice displacements following it nearly adiabatically (cf. Fröhlich 1952). One speaks of a renormalization of various quantities; thus s is the renormalized sound velocity. The elimination of the part of the phonon field following the electron density leads adiabatically to an additional interaction between electrons, as anticipated by our qualitative considerations. As expected, it is attractive when condition (4.1) is satisfied. This means that it strongly depends on the electronic velocities.

4.2. *The Isotope Effect*

To investigate the energy which can be gained by this second type of interaction we notice that adiabatic treatment of the phonon field means that in (3.7) the term containing the rate of change $\dot{\mathbf{P}}$ of the phonon field can be neglected and hence $\mathbf{P}$ can be obtained by finding the minimum potential plus interaction energy for a given electronic configuration. If $\sigma(\mathbf{x})$ is that part of the electron density $n(\mathbf{x})$ which satisfies (4.1), then the displacement $\mathbf{P}(\mathbf{x})$ following $\sigma(\mathbf{x})$ adiabatically is thus obtained by minimizing the relevant part of $H_f + H_{int}$, i.e. putting

$$\delta U = 0, \quad U = \int \{\tfrac{1}{2} M n_i s^2 (\operatorname{div} \mathbf{P})^2 + C\sigma(\mathbf{x}) \operatorname{div} \mathbf{P}\}\, d^3x \qquad \ldots\ldots(4.2)$$

so that

$$\operatorname{div} \mathbf{P}(\mathbf{x}) = -\frac{C\sigma(\mathbf{x})}{M n_i s^2}. \qquad \ldots\ldots(4.3)$$

Hence the relevant part U in the Hamiltonian depending on this type of interaction is obtained by reintroducing (4.3) into U,

i.e.

$$U = -\frac{1}{2}\frac{C^2}{M s^2 n_i} \int \sigma^2(\mathbf{x})\, d^3x. \qquad \ldots\ldots(4.4)$$

Now the density of an assembly of free electrons in their ground state filling up all levels to the Fermi level ζ does not contain any density fluctuations $\sigma(\mathbf{x})$ of the type required in (4.4). To obtain a $\sigma(\mathbf{x}) \neq 0$ it is necessary to transfer some electrons into excited states, and condition (4.1) shows that this should be done in a range less than $\hbar w s$ where w is a representative phonon wave number. The density $\sigma(\mathbf{x})$ obtained in such a manner can thus involve at most a fraction $\hbar w s/\zeta$ of all electrons. Hence

$$\int \sigma^2(\mathbf{x})\, d^3x < \frac{N_{el}^2}{V}\left(\frac{\hbar w s}{\zeta}\right)^2 = \frac{N_{el}^2}{V}\left(\frac{2m\hbar w s}{\hbar^2 k_0^2}\right)^2 \simeq \frac{N_{el}^2}{V}\frac{m s^2}{\zeta}. \qquad \ldots\ldots(4.5)$$

Here k_0 is defined by $\zeta = \hbar^2 k_0^2/2m$, and $k_0 \simeq w$ holds; V is the total volume. Thus if the total number of electrons N_{el} is considered of the same order as the number of ions $n_i V$, then the magnitude of U is less than that of an energy U_0,

$$|U| < |U_0|, \quad U_0 \simeq -\frac{C^2}{\zeta}\frac{m}{M} N_{el}. \qquad \ldots\ldots(4.6)$$

The actual magnitude of U may be considerably less than U_0 but it will differ from U_0 only by a numerical factor. The important result is the term m/M, showing that U is inversely proportional to the isotopic mass M. This, of course, is a direct consequence of the condition that the lattice displacement must follow the electrons in an adiabatic manner, a condition which involves m/M.

The above considerations cannot give a picture of the type of correlation between the electron density and the lattice displacement which, as contemplated above, should lead to a second-order phase transition and to other properties of superconductors. They simply show that if electron–phonon interaction leads to these effects then an energy of the form of U_0 must be involved, i.e. an energy inversely proportional to the isotope mass M. It is useful to write U_0 in the form

$$U_0 = -Fms^2 N_{el}, \quad F = C^2/Ms^2\zeta, \qquad \ldots\ldots(4.7)$$

because the dimensionless parameter F can be determined empirically, for C^2 enters the high-temperature conductivity. F is found to be of order unity, and ms^2 is of the order 10^{-5}–10^{-6} ev. Thus $|U_0|$ approaches the order of magnitude required empirically in the superconductive transition. From an order of magnitude point of view we notice that the principal quantities referring to electrons alone, namely ζ, C or the average Coulomb interaction between electrons, are all of the order of several electron volts. This large magnitude presented one of the main difficulties in finding a starting point for a theory of the superconductive state. It is the possibility of using the phonon velocity s in combination with the electronic mass which provides the chance of a correct order of magnitude. This possibility was, in fact, one of the considerations which led me to contemplate electron–phonon interaction as essential for superconductivity.

Finally, it should be mentioned that the above considerations concerning U are incomplete because a certain positive energy (including some Coulomb energy) is required to establish the required electron density $\sigma(\mathbf{x})$. This does not alter the form of U but imposes a condition on the interaction constant C, requiring C^2 to be sufficiently large. This means that not all metals will exhibit the required interaction, and since C^2 is proportional to the high-temperature resistivity, bad conductors are more likely to become superconductors at low temperatures than good conductors, a result which agrees with experimental evidence (cf. Matthias 1957).

4.3. *The Gap, One-dimensional*

Perturbation theory, though it leads to the isotope effect, does not permit derivation of the correlation between electron densities and lattice displacements which ultimately should lead to the gap.

The difficulties which were encountered in this respect led me to consider a very simplified and unrealistic model: a one-dimensional case with extremely

strong electron–phonon interaction. This model (Fröhlich 1954) provides a gap in a very simple manner. We remember that Peierls has shown in 1930 that a sinusoidal potential of wave number w_0 (acting on electrons) yields in a one-dimensional case a gap in the energy spectrum at electronic wave number $k_0 = w_0/2$. If all states up to the gap are occupied and all states above it are empty, then the electron density has periodic fluctuations with wave number $w_0 = 2k_0$. Now assume in our one-dimensional case a number of electrons such that in their lowest energy state (free electrons) they would occupy all levels up to k_0. Assume further that by electron–phonon interaction the lattice displacement with wave number $w_0 = 2k_0$ is strongly excited. This will change the electron states such as to cause a gap on top of the electron distribution, and also an electron density fluctuation with wave number w_0. This latter in turn causes a lattice displacement w_0. The electron density and the lattice displacement thus maintain each other and moreover lead to a decrease in total energy. If the electron–phonon interaction is very strong, then the gap is very wide, and the influence of phonons of wave numbers different from w_0 can be shown to be very small.

At first sight one might think that an insulator has been created. However, in an infinitely long or in a circular model it can be seen that the periodic density fluctuations of the electrons and of the lattice are only fixed relative to each other, but not relative to the lattice points. The whole configuration can, in fact, move through the lattice. Furthermore, with increasing temperature, electrons are excited across the gap. These electrons reduce the electron density fluctuations which in turn reduce the gap width. This in turn facilitates excitation of electrons across the gap. A situation typical for a second-order phase transition has arisen (cf. also Kuper 1955).

This model thus shows in a simple manner how a gap can arise by electron–phonon interaction. It assumes, however, an enormous strength of interaction such that the dynamical properties of this interaction, expressed by (4.1), are suppressed and all electronic states—not only those selected by (4.1)—contribute to the density which is followed adiabatically by the lattice displacements. As a result, the isotope effect is absent. Nevertheless, the gain of energy U_1, owing to the electron–phonon interaction, has an extremely important mathematical form,

$$U_1 = -\zeta N_{el}\, e^{-1/F}, \quad W = \zeta\, e^{-1/2F} \qquad \ldots\ldots(4.8)$$

where W is the width of the gap, and the dimensionless interaction constant F given by (4.7) is proportional to the square of the constant C entering H_{int}. Perturbation theory up to any order would present the perturbed energy as a power series in C^2, i.e. in F. The result (4.8), however, cannot be developed in such a power series because it has an essential singularity at $F = 0$. This shows that in our one-dimensional model the gap could never have been derived by perturbation theory, a very important conclusion.

It would be desirable now to include the dynamic behaviour resulting in the isotope effect in the present model. The present method unfortunately does not lend itself to this aim. We can conjecture, however, the energy U in this case if we remember that it should differ from U_0 (Eqn (4.7)) by a numerical factor only. For if we assume that the gap is necessarily connected with the

factor $\exp(-1/F)$, then we should expect this factor to replace the factor F in (4.7), i.e.

$$U \simeq -N_{el}\, ms^2\, e^{-1/F}. \qquad \ldots\ldots (4.9)$$

This, indeed, is the expression derived by the new methods described below. It is very sensitive to F and its numerical value is always below that of $|U_0|$, as required. The magnitude of F can be varied experimentally by putting a superconductor under high pressure. The very strong pressure dependence found experimentally is in accordance with (4.9), though no detailed calculation has been made so far.

4.4. *Pairs and the Gap*

In order to develop a new method, Schafroth (1954) has suggested consideration of correlations between electron pairs (this has also been proposed by Peierls (1955)). He pointed out (cf. also Schafroth, Butler and Blatt 1957) that an electron pair above, but near the top, of the Fermi surface (i.e. near the energy ζ) may form a pseudo-molecule, i.e. the mean distance between the two electrons will remain finite, provided an attractive interaction exists. They are bound (in space) although they have positive energy. The reason for this behaviour is understandable if one assumes that the attractive interaction does not act between electrons in fully occupied levels, i.e. does not affect the electrons up to ζ. Two additional electrons slightly above ζ, in view of their attraction, will then have a resultant energy below 2ζ, say $2(\zeta-\lambda)$. If they were to disperse outside the range of the attractive force, each would have to go into an energy level $\zeta-\lambda$. This is impossible, in view of the Pauli exclusion principle, because these states are occupied. Cooper (1956) has made a quantitative calculation of the energy λ on the basis of the electron–phonon interaction. He found $\lambda \simeq (ms^2\zeta)^{1/2} \exp(-1/F)$, clearly a highly satisfactory result in view of (4.9).

The value of the above argument is doubtful because no interaction will be completely absent for fully occupied states (cf. also § 5). Nevertheless, the idea of pair correlation has inspired Bardeen, Cooper and Schrieffer (1957) to propose a state function which for the first time led to a realistic derivation of the gap and of an energy U, in agreement with the conjectured value (4.9). To obtain this result these authors had to make some rather bold assumptions concerning the electron interaction induced by the phonon field. Fortunately, however, their results were confirmed by a mathematically more satisfactory method, applied soon afterwards by Bogoliubov (1958 a, b ; cf. also Valatin 1958) forming a generalization of a method which this author had developed already in 1947 for other purposes. The new method permitted calculation of many detailed properties of superconductors (for Meissner effect and superconductivity see, however, § 5), and good agreement with experiments was obtained (cf. Bardeen and Schrieffer 1961).

Following the general trend of the present article, our next step should have been a qualitative description of the physical reasons for the existence and magnitude of the gap, and its connection with pair correlation. Unfortunately, this seems to be impossible at the present moment. The reason for this may be found in a lack in my personal ability ; I believe, however, that it might also be found in a defect of the method itself which, although very elegant, contains features which evade interpretation.

§ 5. Criticisms; Meissner Effect

The derivation of the gap by Bardeen, Cooper and Schrieffer (1957), the justification of their results and the development of a new method by Bogoliubov (1958a, b; reviewed by Bogoliubov, Tolmachev and Shirkov 1959) led to a great mathematical activity, for this method could be applied to the calculation of many detailed features, depending on the existence of the gap. Yet, as pointed out at the end of § 4, it has so far been impossible to understand the existence of the gap in terms of a simple physical picture. The present article does not deal with mathematical developments; in order to understand the difficulties which one meets in the new method it is necessary, however, to go briefly into some mathematical details. For this purpose we develop the wave functions $\psi(\mathbf{x}), \psi^*(\mathbf{x})$ of the electron field used, e.g. in (2.12) and (3.5), into a spatial Fourier series (s is the spin quantum number which can have two values, $\uparrow$ and $\downarrow$; $\xi(s)$ are spin functions),

$$\psi(\mathbf{x}) = \frac{1}{V^{1/2}} \sum_{\mathbf{k},s} a_{\mathbf{k},s}\, \xi(s)\, e^{i\mathbf{k}\mathbf{x}}, \quad \psi^*(\mathbf{x}) = \frac{1}{V^{1/2}} \sum_{\mathbf{k},s} a_{\mathbf{k},s}{}^*\, \xi^*(s)\, e^{-i\mathbf{k}\mathbf{x}}, \qquad \ldots\ldots(5.1)$$

so that the electron density $n(\mathbf{x})$, for instance, is given by (cf. (3.6); the integration is carried out over spin as well as space variables)

$$n(\mathbf{x}) = \frac{1}{V} \sum_{\mathbf{k},s} \sum_{\mathbf{k}',s'} a_{\mathbf{k},s}{}^*\, a_{\mathbf{k}',s'}\, \xi^*(s)\, \xi(s') \exp\left[i(\mathbf{k}'-\mathbf{k})\mathbf{x}\right], \quad N_{\mathrm{el}} = \sum_{\mathbf{k},s} a_{\mathbf{k},s}{}^*\, a_{\mathbf{k},s}. \qquad \ldots\ldots(5.2)$$

Here the $a_{\mathbf{k},s}, a_{\mathbf{k},s}{}^*$ are not numbers but non-commuting operators such that $a_{\mathbf{k},s}{}^*\, a_{\mathbf{k},s}$ can have the eigenvalues 0 and 1 only as required by the Pauli principle. The state of the system of electrons is described by a state vector Ψ. It has the property that $a_{\mathbf{k},s}\Psi$ is zero when Ψ describes a state in which the level $\mathbf{k}, s$ is unoccupied; when this state is occupied, however, then $a_{\mathbf{k},s}\Psi$ is equal to a new state vector in which $\mathbf{k}, s$ is empty, so that $a_{\mathbf{k},s}{}^2\Psi$ is always zero. Similarly, $a_{\mathbf{k},s}{}^*\Psi$ is zero when $\mathbf{k}, s$ is already occupied; when unoccupied, $a_{\mathbf{k},s}{}^*\Psi$ represents a state vector in which $\mathbf{k}, s$ is occupied. $a_{\mathbf{k},s}{}^*$ and $a_{\mathbf{k},s}$ are denoted as creation and annihilation operators respectively. Now Bardeen, Cooper and Schrieffer (1957) propose for the ground state a state vector of the form

$$\Psi = \prod_{\mathbf{k}} \left(u_{\mathbf{k}} + v_{\mathbf{k}}\, a_{\mathbf{k}\uparrow}{}^*\, a_{-\mathbf{k}\downarrow}{}^*\right) \Psi_0, \qquad \ldots\ldots(5.3)$$

where the $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ are numerical constants satisfying

$$u_{\mathbf{k}}{}^2 + v_{\mathbf{k}}{}^2 = 1, \qquad \ldots\ldots(5.4)$$

and Ψ_0 is the state vector of the vacuum.

Equation (5.3) is equivalent to the assumption that two levels $\mathbf{k}, -\mathbf{k}$ with opposite momentum are either both empty, or they are both occupied (with electrons with antiparallel spin). We notice that a pair of electrons (with antiparallel spins) whose centre of gravity is at rest satisfies a similar condition; its state vector has the form

$$\Phi\Psi_0 = \sum_{\mathbf{k}} c_{\mathbf{k}}\, a_{\mathbf{k}\uparrow}{}^*\, a_{-\mathbf{k}\downarrow}{}^*\, \Psi_0, \qquad \ldots\ldots(5.5)$$

if the $c_{\mathbf{k}}$ are numerical constants. If we imagine the product in (5.3) to be carried out, then a sum is obtained whose terms contain products of various pairs $a_{\mathbf{k}\uparrow}{}^* a_{-\mathbf{k}\downarrow}{}^*$, each creating a pair of electrons such that the Pauli principle holds. The state described by Ψ does not, therefore, refer to a fixed number of electrons, but the mean number fluctuation around the average is found to be small provided this average number N_{el} of electrons is large. Blatt (1960) has shown that if we select from Ψ the terms referring to the definite number N_{el} of electrons, then the constants $c_{\mathbf{k}}$ can be chosen such that

$$\Psi = \Phi^{N_{\mathrm{el}}/2}\Psi_0. \qquad \ldots\ldots(5.6)$$

Thus Ψ can be expressed in terms of pair functions.

Calculation of the energy of the ground state can now be performed with the help of the Hamiltonian by an appropriate choice of the $u_{\mathbf{k}}, v_{\mathbf{k}}$. Excited states, and hence the gap, follow in a similar manner. Such calculations require various approximations which are not always clearly definable.

The Bogoliubov method in the first place replaces the a, a^* by other operators α, α^*,

$$\alpha_{\mathbf{k},0} = u_{\mathbf{k}} a_{\mathbf{k}\uparrow} - v_{\mathbf{k}} a_{-\mathbf{k}\downarrow}{}^*, \quad \alpha_{\mathbf{k},1} = u_{\mathbf{k}} a_{-\mathbf{k}\downarrow} + v_{\mathbf{k}} a_{\mathbf{k}\uparrow}{}^*, \qquad \ldots\ldots(5.7)$$

and similarly for the α^*. Clearly, the a^*, a can be uniquely expressed by the α^*, α, and hence the Hamiltonian too can be expressed in terms of these new operators.

As a consequence, the Hamiltonian can be written as a sum

$$H = H_0 + H_1 \qquad \ldots\ldots(5.8)$$

where H_0 contains the α, α^* only by way of a sum over terms proportional to $\alpha_{\mathbf{k}}{}^* \alpha_{\mathbf{k}}$. H_1, on the other hand, depends on the α in a more complicated manner. It is assumed then that the $u_{\mathbf{k}}, v_{\mathbf{k}}$ can be chosen such that H_0 forms a good approximation to H. For this purpose, the $u_{\mathbf{k}}, v_{\mathbf{k}}$ are determined from a condition which essentially means that treatment of H_1 by perturbation theory shall lead in lowest order to the smallest possible correction to the energy. State vectors solving (diagonalizing) the Hamiltonian H_0 are obtained in a very simple way in terms of the operators α and α^*. They show the existence of a gap and in fact lead to the same results as those obtained by the method based on (5.3). The new method, however, shows clearly which terms in the Hamiltonian (i.e. H_1) have been neglected.

5.1. *Gauge Invariance*

In order to discuss some of the difficulties of the methods outlined above, it is necessary to introduce the concept of gauge invariance. Consider an electromagnetic field $\mathbf{E}, \mathbf{H}$, which (cf. also (2.7)) may always be derived from a vector potential $\mathbf{A}$, and a scalar potential S by

$$\mathbf{H} = \operatorname{curl} \mathbf{A}, \quad \mathbf{E} = -\frac{1}{c}\dot{\mathbf{A}} + \operatorname{grad} S. \qquad \ldots\ldots(5.9)$$

Consider also a pair of electrons with coordinates $\mathbf{x}_1$ and $\mathbf{x}_2$ interacting through $V(\mathbf{x}_1-\mathbf{x}_2)$. They satisfy a Schrödinger equation

$$-\frac{\hbar^2}{2m}\left\{\left[\frac{\partial}{\partial \mathbf{x}_1}-i\frac{e}{\hbar c}\mathbf{A}(\mathbf{x}_1)\right]^2+\left[\frac{\partial}{\partial \mathbf{x}_2}-i\frac{e}{\hbar c}\mathbf{A}(\mathbf{x}_2)\right]^2\right\}+V(\mathbf{x}_1-\mathbf{x}_2)+eS(\mathbf{x}_1)+eS(\mathbf{x}_2),$$

$$\phi(\mathbf{x}_1,\mathbf{x}_2,t)=\frac{\hbar}{i}\frac{\partial}{\partial t}\phi(\mathbf{x}_1,\mathbf{x}_2,t). \qquad \ldots\ldots(5.10)$$

Clearly the electromagnetic field (5.9) remains unaltered if we 're-gauge' $\mathbf{A}$ and S by making the replacements

$$\mathbf{A}\to\mathbf{A}+\operatorname{grad}\chi, \quad S\to S+\frac{1}{c}\dot{\chi}, \qquad \ldots\ldots(5.11)$$

where χ is a space time function which must satisfy some very simple requirements derived in electrodynamics. We notice that the replacements (5.11) require a change of the phase of the Schrödinger function by

$$\phi\to\phi\exp\{i(e/\hbar c)[\chi(\mathbf{x}_1,t)+\chi(\mathbf{x}_2,t)]\}. \qquad \ldots\ldots(5.12)$$

The above procedure can be generalized to any larger number of electrons, or to an electron field. It is always possible to re-gauge the potentials and the state vectors by a procedure similar to (5.11) and (5.12). Frequently it is believed that the possibility of such gauge transformations is necessarily connected with the presence of an electromagnetic field. This is not so, for by putting

$$\mathbf{A}=\operatorname{grad}\chi, \quad S=\frac{1}{c}\dot{\chi} \qquad \ldots\ldots(5.13)$$

we ensure by (5.9) that the electromagnetic field vanishes and yet (5.12) holds.

One of the criticisms to be levelled against formulae (5.3), (5.5) and (5.7) is that they have not been formulated in a gauge-invariant manner even in the absence of an electromagnetic field. This also implies that the approximate Hamiltonian H_0 introduced in (5.8) is not gauge-invariant. Since the original Hamiltonian is gauge-invariant, it follows that H_1 must correct the lack of gauge invariance of H_0. Now it was decided to treat H_1 by perturbation theory. There is no limit, however, to the magnitude of the function χ nor to its rate of spatial and temporal change. Thus χ may have features which make treatment of H_1 by perturbation theory a very doubtful procedure. Yet χ has no physical meaning whatsoever and hence should not be able to influence the significance of solutions. I believe that the difficulty in arriving at a simple physical picture for the manner in which the gap is established is connected with this difficulty of introducing a gauge function χ. For, as a consequence, already the pair function Φ (Eqn (5.5)) does not permit a simple physical interpretation. This is closely connected with the fact that its centre of gravity is always assumed to be at rest if $\chi=0$.

It is possible, of course, to work with pair functions which are gauge-invariant, as has been shown by Bogoliubov *et al.* (1959) and by Blatt (1960). For this purpose we notice that (5.5) corresponds in particle coordinates to a function of the relative coordinates $\mathbf{x}_1-\mathbf{x}_2$ only while the transformation (5.12) requires a more general function of $\mathbf{x}_1$ and $\mathbf{x}_2$. Thus a gauge-invariant pair function cannot

have the form (5.5) (centre of gravity necessarily at rest) but must be of the more general form

$$\Phi_g = \sum_{k,s} \sum_{k',s'} c_{ks,k's'} a_{ks}^* a_{k's'}^*, \qquad \ldots\ldots(5.14)$$

which does permit motion of the centre of gravity. One might try, therefore, to replace in (5.6) the pair function Φ by Φ_g. Unfortunately, however, Φ_g is so general that a specific solution of our problem has not been carried out yet. In fact, I feel that a physical idea is required which would lead to a more specific form of the constants $c_{ks,k's'}$.

In this connection it is of relevance to say a word about so-called collective excitations. It will be remembered that solutions based on H_0 yield the ground state, and the gap. In terms of electron pairs (pseudo molecules) this means that excited states exist in which the pair (or several pairs) are broken up, the dissociation energy corresponding to the gap width. Use of H_1 leads to other types of excitations corresponding to a motion of the centre of gravity of pseudo molecules, as shown by Bogoliubov, Tolmachev and Shirkov (1959) and others. As has been shown above, the possibility of such centre-of-gravity motion is required already in the ground state if gauge invariance is satisfied. Since $H_0 + H_1$ but not H_1 is gauge-invariant, it is obvious that H_1 should induce this type of motion. It is dangerous, however, without further discussion, to interpret such terms as excitations, for already the ground state if described in a gauge-invariant manner must contain such terms.

5.2. *Other Criticisms*

The method described in connection with Eqns (5.1)–(5.8) can be applied to Hamiltonians different from (3.4). It leads to a gap whenever the particles it refers to satisfy the Pauli principle and have a weak attractive interaction. I feel that such a conclusion may well be incorrect and that possibly a strong velocity dependent interaction only—like that obtained through the phonon field—leads to a gap. Such an interaction is restricted to particles whose energy difference is relatively small,† as indicated by (4.1).

This question is closely connected with the formation of bound pairs (pseudo molecules) near the Fermi surface, as discussed in § 4. As has been mentioned, the argument given there is open to criticism for it is wrong to assume that the (attractive) interaction is completely absent between electrons in fully occupied states. It follows that the selected two electrons outside the Fermi surface prevent (virtual) transitions of the electrons inside the surface to the states occupied by these two electrons, just as these two electrons cannot jump into the interior. In order to decide whether or not the two extra electrons form a bound state it is necessary, therefore, to calculate the change of energy of the whole system, i.e. to include the change of energy of all inner electrons by virtue of their reduced chance of making virtual transitions. One would think that in case of an interaction which acts only between electrons of nearly equal energy the conditions for

† Applications of the method have been made to show that a gap exists in the (incomplete) outer shell of a heavy nucleus. Such applications consider interactions between the nucleons in the outer shell only. They ignore interactions with nucleons in inner shells. This corresponds to a restriction of the relevant interaction to a relatively narrow energy region, similar to (4.1).

the formation of bound pairs are easier to fulfil than for a purely distance-dependent interaction.

The question whether only a particular type of (attractive) interaction can produce a gap is related to the question of the convergence of the perturbation series due to H_1 in (5.8). Such convergence is extremely difficult to prove. It should be remembered in this connection that H_0 yields a gap for the ground state, but that states without a gap are very close to it energetically. It is feasible that for some types of interaction the Hamiltonian H_1 mixes these gapless states into a (higher order) ground state which then might no longer exhibit a gap. Perturbation theory (for H_1) would yield this only in a very high order (proportional to the total number of electrons).

5.3. *Meissner Effect*

Though superconductivity (persistent currents) is perhaps more spectacular, the Meissner effect is the more basic electromagnetic property as it refers to a static, rather than to a stationary case. The phenomenological discussion in § 2 has shown that in order to derive this effect one must simply show that a magnetic field in first order does not change the state of the system. The gap model was seen to lead to such a result provided the matrix elements for transitions across the gap are not changed by the setting-up of the gap, but this assumption was found to contradict the uncertainty relation. This condition on the matrix elements may, of course, be modified by requiring that the inevitable change of these matrix elements by the setting-up of the gap is not very substantial.

The solutions based on H_0 discussed above yield a Meissner effect because it is found indeed that the appropriate matrix elements are only insignificantly changed by the setting-up of the gap (Bardeen, Cooper and Schrieffer 1957). This result is open to objections, however, in view of the lack of gauge invariance (for a discussion see Schafroth 1960). One may try to restore gauge invariance by including H_1 in terms of perturbation theory, for as discussed above, the total Hamiltonian H_0+H_1 is gauge-invariant. This procedure confirms the existence of the Meissner effect as shown, e.g. by Rickayzen (1959). Such a derivation of the Meissner effect is, therefore, based on the assumption that the perturbation series based on H_1 converges. Do we accept this as a valid derivation? I think that, at present, the only answer to this question can be that we accept it if we show goodwill. The question of providing a simple physical picture, however, remains unanswered, just as in the case of the gap.

Another line of thought (cf. Schafroth 1960) makes more explicit use of the existence of pseudo molecules. These latter, if they could be treated as separate entities, would represent a gas satisfying Bose statistics. Now an ideal Bose gas shows (i) a phase transition into a 'condensed' state, and (ii) it shows a Meissner effect in this state. It is hoped in this line of thought that our pseudo molecules still show these features, although they certainly cannot be treated as independent of each other. This idea would provide a simple physical picture for both gap and Meissner effect. So far, however, it has not been possible to prove the required conditions.

Finally, it should be mentioned that no serious attempts have been made yet to derive the existence of persistent currents.

§6. The Psychology of Superconductivity

The development of a satisfactory theory of the superconductive state was connected with so much personal emotion that an article describing this development would be incomplete without mentioning this. Necessarily this section expresses my personal view, since editors of scientific journals are very careful to eliminate from papers any reference to such questions; they usually succeed in making scientific papers extremely dull. I should like to take this opportunity for proposing the establishment of a new journal. It should contain discussions including criticisms of questions of current interest, and of papers, books and articles.

To return to superconductivity. By 1933, when it was clear that the basic physics required for the development of a theory of superconductivity was well understood, many physicists had tried in vain to propose such a theory. A general feeling of frustration arose such that the first thing one did when acquainted with another attempt was to try to find where it was wrong. This feeling found its expression in Pauli's remark, "Theories of superconductivity are wrong", and by Felix Bloch's theorem, "Theories of superconductivity can be disproved". It was astonishing to see how my first paper in 1950—soon followed by Bardeen's first paper—led to a complete reversal of opinion. Schafroth (1951) followed Pauli's suggestion to work further on this theory and he was the first to show that perturbation theory could not be applied. Yet neither he, nor most others, doubted the correctness of the basic assumption that electron–phonon interaction should be principally responsible for superconductivity. This, so far as I have seen, was principally due to the fact that no 'new' interaction between electrons had to be invented but that the theory was completely based on the empirically successful free electron model. Discovery of the isotope effect was, of course, an important feature in this change of opinion. Yet it appears that all this did not affect some of those physicists who had previously formed an opinion in a different direction. A striking example can be found in the article by W. L. Ginsburg published in 1953. This author, it appears, had formed the opinion that the free electron model can never succeed in accounting for the properties of superconductors but that the electrostatic interaction between electrons—modified by the high dielectric constant of metals—should be of primary importance. He must have been very uncertain of the correctness of his opinion and yet unprepared to give it up, for otherwise it would be difficult to understand the angry and quite incorrect report he gives on the early papers† dealing with electron–phonon interaction. Also it is not without amusement that one reads now his dogmatic prescriptions at the end of his article as to how work should proceed in order to produce a correct theory. At the same time one wonders whether this author's influence was not in some way responsible for the fact that Bogoliubov did not apply his successful method earlier.

In contrast to this dangerous dogmatic approach, the development of the theory of superconductivity is an excellent example of the intuitive approach. My idea was foremost that one should consider the small energy involved in the

† The editors of the journal had asked me at the time to comment on this article; I suggested, however, that this should be unnecessary as the development of the following years no doubt would (as they did) justify the approach.

superconductive transition as the first problem to be answered. Electron–phonon interaction provided this possibility from a purely dimensional point of view. When the early attempts did not succeed in solving the problem, then it was required, I thought, to find a new approach to the many-body problem involved. However, intuition won again: Bardeen, Cooper and Schrieffer (1957) proposed their state vector (5.3) which virtually solved the problem of the gap.

The great activity which arose from this paper and from the use of Bogoliubov's (1958 a, b) method led to many successful applications. In view of this success, not every worker in the field agrees with the criticism expressed in § 5. Such difference of opinion, I think, might be stimulating to all concerned. Unfortunately, it appears that in recent years theoretical work on superconductivity has frequently led to concern with keeping the upper hand in a discussion rather than with illuminating its subject. An example of this occurs in an article in *Physics Today*, Vol. 13, No. 6, p. 34, reviewing recent work published in *Soviet Physics*. After reference to work not comparable in importance with that of Bogoliubov, there occurs the statement: "Another group that carries on theoretical studies in superconductivity is that of Bogoliubov.... Publications of this group are frequently characterized by their failure to mention the B.C.S. (= Bardeen, Cooper and Schrieffer) theory or the work of physicists at the Institute for Physical Problems...."

§ 7. Outlook

The present article shows that the theory of the superconductive state has reached a stage now in which it resembles the theory of many other processes in solid state physics. With some goodwill one can say that the theory accounts for the principal observed effects. In particular, we know that the electron–phonon interaction which at higher temperatures is responsible for ordinary conductivity can, at very low temperatures, lead to an extremely small energy change connected with the setting-up of an energy gap and with a violent change in most electronic properties. A number of features are not yet well understood; superconductivity has this in common with many other many-body problems. Thus the reader should be reminded that no simple and clear theory of melting does exist yet.

I believe that the many-body problem still contains some secrets which require deep thought to be revealed. In this connection I wish to recall that the superfluidity of liquid helium II is a property which shares many features with superconductivity, the most important being probably the existence of an energy gap for certain excitations. If this is a very general property of fluids at low temperature (in superconductors we might speak of an electron fluid) then it might be of interest to resume work on quantum hydrodynamics. The idea has been expressed some time ago (Landau 1941, 1947) that rotons with finite rest energy should exist in quantum fluids, i.e. that a gap should exist for the excitation of transverse motion. A careful investigation by Allcock and Kuper (1955), following earlier work by Ziman (1953), has shown that quantization of the transverse motion of a continuous fluid leads to certain difficulties which can be overcome only by introducing an atomistic structure. This means that in quantum mechanics the usual hydrodynamics does not represent a consistent theory. I have a feeling that without going into a detailed atomistic theory the above difficulties might be overcome by

the introduction of non-local features into hydrodynamics. From this modification conditions might possibly arise under which quantization would lead to a finite energy for the excitation of transverse motion. The similarity of substances so different as liquid helium and the electron fluid of superconductors certainly points to some very general features which should be closely connected with quantization in macroscopic physics (cf. F. London 1950). This possibility may help, perhaps, to restore some fascination to the subject of this article.

References

Allcock, G. R., and Kuper, C. G., 1955, *Proc. Roy. Soc.* A, **231**, 226.

Allen, W. D., Dawton, R. H., Bär, M., Mendelssohn, K., and Olsen, J. L., 1950 a, *Nature, Lond.*, **166**, 1071.

Allen, W. D., Dawton, R. H., Lock, J. M., Pippard, A. B., and Schoenberg, D., 1950 b, *Nature, Lond.*, **166**, 1071.

Bardeen, J., 1950, *Phys. Rev.*, **80**, 567 ; 1951, *Rev. Mod. Phys.*, **23**, 261 ; 1956, *Handb. d. Phys.*, **15**, 274 (Berlin : Springer).

Bardeen, J., Cooper, L. N., and Schrieffer, J. R., 1957, *Phys. Rev.*, **106**, 162 ; **108**, 1175.

Bardeen, J., and Schrieffer, J. R., 1961, *Progress in Low Temperature Physics*, **3**, in the press.

Becker, R., Sauter, F., and Heller, G., 1933, *Z. Phys.*, **85**, 772.

Blatt, J. M., 1960, *Progr. Theor. Phys.*, **23**, 447.

Bloch, F., 1928, *Z. Phys.*, **52**, 555.

Bogoliubov, N. N., 1947, *J. Phys. Moscow*, **9**, 23 ; 1958 a, *J. Exp. Theor. Phys.*, **34**, 58, 173 (*Sov. Phys.*, **7**, 41, 51) ; 1958 b, *Nuovo Cim.*, **7**, 794.

Bogoliubov, N. N., Tolmachev, V. V., and Shirkov, D. V., 1959, *A New Method in the Theory of Superconductivity* (New York : Consultants Bureau).

Bohm, D., and Pines, D., 1953, *Phys. Rev.*, **92**, 609.

Buckingham, M. J., 1957, *Nuovo Cim.*, **5**, 1763.

Cooper, L. N., 1956, *Phys. Rev.*, **104**, 1189.

Corak, W. S., Goodman, B. B., Satterthwaite, C. B., and Wexler, A., 1956, *Phys. Rev.*, **102**, 656.

Ehrenfest, P., 1933, *Proc. Acad. Amsterdam*, **36**, 153.

Fröhlich, H., 1950, *Phys. Rev.*, **79**, 845 ; *Proc. Phys. Soc.* A, **63**, 778 ; 1952, *Proc. Roy. Soc.* A, **215**, 291 ; 1953, *Physica*, **19**, 755 ; 1954, *Proc. Roy. Soc.* A, **223**, 296.

Fröhlich, H., Pelzer, H., and Zienau, S., 1950, *Phil. Mag.*, **41**, 221.

Ginsburg, W. L., 1953, *Fortschr. Phys.*, **1**, 101.

Glover, R. E., and Tinkham, M., 1956, *Phys. Rev.*, **104**, 844 ; 1958, *Phys. Rev.*, **108**, 243.

Goodman, B. B., 1953, *Proc. Phys. Soc.* A, **66**, 217.

Gorter, C. J., 1933, *Archiv Teyler*, **7**, 387.

Gorter, C. J., and Casimir, H. G. B., 1934, *Phys. Z.*, **35**, 963.

Kamerlingh-Onnes, H., 1911, *Leiden Comm.*, **122b**, **124c**.

Keesom, W. H., 1924, *Fourth Solvay Congress*, 289.

Kuper, C. G., 1955, *Proc. Roy. Soc.* A, **227**, 214 ; 1959, *Advanc. Phys.*, **8**, 1.

Landau, L. D., 1933, *Z. Phys. Sowjet*, **3**, 664 ; 1941, *J. Phys. Moscow*, **5**, 71 ; 1947, *J. Phys. Moscow*, **11**, 91.

London, F., 1950, *Superfluids* (New York : John Wiley).

London, F., and London, H., 1935, *Physica*, **2**, 341.

Matthias, B., 1957, *Progress in Low Temperature Physics*, **2**, 138.

Maxwell, E., 1950, *Phys. Rev.*, **78**, 477.

Meissner, W., and Ochsenfeld, R., 1933, *Naturwissenschaften*, **21**, 787.

Peierls, R., 1930, *Ann. Phys., Lpz.*, **4**, 121; 1955, *Quantum Theory of Solids* (Oxford: University Press), p. 220.

Pippard, A. B., 1953, *Proc. Roy. Soc.* A, **216**, 547.

Reynolds, C. A., Serin, B., Wright, W. H., and Nesbitt, L. B., 1950, *Phys. Rev.*, **78**, 487.

RICKAYZEN, G., 1959, *Phys. Rev.*, **115**, 795.
SCHAFROTH, M. R., 1951, *Helv. Phys. Acta*, **24**, 645; 1954, *Phys. Rev.*, **96**, 1442; 1960, *Solid State Physics*, **10**, 293.
SCHAFROTH, M. R., BUTLER, S. T., and BLATT, J. M., 1957, *Helv. Phys. Acta*, **30**, 93.
SHOENBERG, D., 1952, *Superconductivity*, 2nd edn (Cambridge: University Press).
SOMMERFELD, A., 1928, *Z. Phys.*, **47**, 1.
VALATIN, J. G., 1958, *Nuovo Cim.*, **7**, 843.
WELKER, H., 1938, *Bayerische Akademie der Wissenschaften*, 115.
ZIMAN, J. M., 1953, *Proc. Roy. Soc.* A, **219**, 257.

HIGH CURRENT GAS DISCHARGES†

By A. A. WARE‡

John Jay Hopkins Laboratory for Pure and Applied Science, General Atomic Division of General Dynamics Corporation, San Diego, California

CONTENTS

PAGE

§ 1. Introduction ... 25
1.1. The class of discharges to be reviewed ... 25
1.2. Introductory and historical survey ... 26

§ 2. The plasma equations and some physical concepts ... 33
2.1. A set of plasma equations ... 33
2.2. The magnetohydrodynamic approximations ... 36
2.3. The collisionless plasma approximation ... 36
2.4. Simple physical concepts ... 37
2.4.1. Magnetic pressure ... 37
2.4.2. 'Freezing in' of magnetic lines of force ... 38

§ 3. Discharge initiation and contraction ... 38
3.1. Ionization ... 38
3.2. Discharge contraction ... 39
3.2.1. High rate of rise of current, no axial magnetic field 39
3.2.2. High rate of rise of current with an axial magnetic field ... 42
3.2.3. Low rate of rise of current, no axial magnetic field 43
3.2.4. Low rate of rise of current with an axial magnetic field ... 43

§ 4. Discharge equilibrium ... 44
4.1. Linear discharges ... 44
4.2. Toroidal discharges ... 45

§ 5. Magnetohydrodynamic stability ... 46
5.1. Theory ... 46
5.1.1. Perfectly conducting compressible cylinder with infinitely thin current sheath ... 47
5.1.2. Perfectly conducting cylinder with finite current sheath ... 48
5.1.3. Perfectly conducting discharge with constant pitch magnetic field ... 49
5.1.4. Discharges with $B_z \gg B_\theta$ ... 49
5.1.5. Discharge models with dissipative effects ... 50
5.1.6. Stable configurations ... 50
5.1.7. Toroidal discharges ... 50
5.2. Experimental observations on stability ... 50
5.2.1. Discharges with trapped axial magnetic field ... 50
5.2.2. Discharges with approximately constant pitch magnetic fields ... 52
5.2.3. Discharges with $B_z \gg B_\theta$ ... 52
5.3. Instability waves ... 53

† Research on controlled thermonuclear reactions is a joint programme carried out by the General Atomic Division of General Dynamics Corporation and the Texas Atomic Energy Research Foundation.

‡ On leave of absence from Associated Electrical Industries Research Laboratory, Aldermaston, Berks.

Vol. 24
1961

PAGE

§ 6. Particle heating and energy loss 54
6.1. The electron temperature and energy loss 54
6.2. The positive ion temperature and heating mechanism 56

§ 7. Nuclear reactions 57
7.1. The linear pinch without an axial magnetic field 57
7.2. The linear pinch with an axial magnetic field 58
7.3. The toroidal pinch with magnetic field 59

§ 8. High current discharges with other geometrical configurations 59
8.1. Azimuthal discharges 59
8.2. Annular discharges ('triax' and 'hard-core') 61
8.3. Rotating plasmas 62
8.4. Plasma 'guns' 63

Acknowledgments 63

References 64

Abstract. In this review high current discharges are taken to be those discharges whose self-magnetic fields play an important role in their behaviour. Such discharges have been studied mainly in connection with the quest for controlled thermonuclear reactions. Of the many possible configurations for the discharge tube and the electromagnetic fields, the pinch discharges in straight or toroidal tubes have received the most attention and, accordingly, a major part of this review is devoted to these discharges. Other configurations are reviewed in a separate section. Following an historical and introductory review of the subject, a set of appropriate plasma equations is presented, and the common approximations and physical concepts associated with the theory are discussed. The properties of the pinch discharge are considered under the headings—discharge initiation and contraction, equilibrium, magnetohydrodynamic stability, particle heating and energy loss, and nuclear reactions—and in many cases it is necessary to subdivide the sections according to whether the rate of rise of current is high or low.

§ 1. Introduction

1.1. *The Class of Discharges to be Reviewed*

Throughout this article the relative term 'high current gas discharge' will be taken to mean a discharge in which the magnetic field produced by the current itself plays a major role in the mechanism of the discharge. This will be the case when the reaction between the self-magnetic field and the current produces a force on the ionized gas comparable with the other forces acting on it, such as the gradient of pressure. The actual magnitude which a current must have to satisfy this condition will in general be higher the higher the gas density in the discharge, but as a rough guide it can be said that the currents referred to will be greater than about 10^3 A and the gas densities less than about 1 mm Hg at N.T.P.

The impetus to study very high current discharges has come from the fact that they present a possible means of containing and heating a gas to very high temperatures and hence inducing controlled thermonuclear reactions. Most of the research in this field has been done with this ultimate goal in view. However, this paper is not a comprehensive review of thermonuclear research since it deals with only one of the two general methods of approach to produce such reactions.

The systems dealt with here start with a cold gas which is ionized and heated by passing a high current through it. In the other class of devices a magnetic field configuration is created in a vacuum system by means of external coils, and a hot gas is generated by the injection of an ordered beam of energetic particles. Although the ultimate goal for this latter class of systems is the same, at their present state of development they bear little resemblance to gas discharges. Their physical properties and problems are closer to those for particle accelerators. It is therefore logical to omit such systems from the present review. For a general review of the various methods of approach to controlled thermonuclear power and the physics involved, the reader is referred to the excellent accounts of Post (1956), Bishop (1958) and Simon (1959).

The most commonly adopted geometrical configurations for the study of high current discharges have been the straight cylindrical tube with end electrodes, and the torus without electrodes. Most of the sections in this article will deal with such discharge tubes. For many of the discharge properties the curvature of the toroidal tube is neglected and no distinction will be made between the two types of tubes. Since most of the discharge properties satisfy simple scaling laws (Bickerton and London 1958) reference will seldom be made to the tube size in a particular experiment. High current discharges have been studied in other geometrical configurations ; these will be reviewed in § 8.

1.2. *Introductory and Historical Survey*

It has been well known to physicists and electrical engineers for many years that a group of parallel conductors carrying currents in the same direction mutually attract one another because of their magnetic fields. That a similar effect will also occur when a current is passed through an ionized gas was first considered theoretically by Bennett (1934) and later by Tonks (1939). Other early treatments were by Fetz (1949), Alfvén (1950), Schluter (1950), Blackman (1951) and Thonemann and Cowhig (1951). The parallel currents flowing in different parts of the gas attract one another and a cylindrical discharge (with current parallel to the axis) will contract in the radial direction towards its axis. This now well-known 'pinch effect' will continue until the attractive force is balanced by a radial pressure gradient. The force per unit volume on the ionized gas is $\mathbf{j} \wedge \mathbf{B}$, where $\mathbf{j}$ is the current density and $\mathbf{B}$ the magnetic field. In this simple case where $\mathbf{j}$ is parallel to the z axis, and $\mathbf{B}$ is in the θ direction, if interaction with neutral gas can be neglected, the equilibrium is given by

$$j_z B_\theta = -\frac{dp}{dr}. \qquad \ldots\ldots(1)$$

This equation can be combined with Maxwell's equation for the current and integrated (see Pease 1957) giving

$$NkT = \tfrac{1}{2}I^2 + \pi b^2 p_b \qquad \ldots\ldots(2)$$

where N is the number of particles (positive and negative) per unit length of the discharge tube, T their average temperature, p_b the pressure at the wall whose radius is b, and I the total current. Clearly the pinch effect will become important when $\frac{1}{2}I^2$ is comparable with NkT. At high currents the constriction of the

discharge causes p_b to become small and

$$NkT \simeq \tfrac{1}{2}I^2. \qquad \ldots\ldots(3)$$

This equation, often referred to as the Bennett relation, gives the magnitude of current required to 'contain' a given number of particles at a given temperature so that the pressure at the wall will be small.

The first reported demonstration of any pinch effect in a gas discharge was by Fetz (1941), who showed that in a low-pressure mercury arc of about 100 A the expected slight sharpening of the radial density distribution in the positive column did occur. Similar experiments were performed by Thonemann and Cowhig (1951) and Mamyrin (1953). The first person, however, to study a current sufficiently high to be within the present criterion was Smith (1941), who induced currents of 400 A in a torus containing low-pressure mercury vapour. The experiments suggested that the self-magnetic field of the current was not causing complete containment of the ionized gas (Smith 1947). About the same time Steenbeck and Hoffman (1943) constructed a high current toroidal discharge device known as the 'Wirbelrohr' in which current pulses up to 10 000 A were induced in an attempt to produce an electron accelerator. Steenbeck recognized that the self-magnetic field within a high current discharge is ideally suited for the focusing and containment of energetic particles. He was also aware that since the Coulomb scattering cross section for charged particles decreases inversely as the square of the particle energy, some of the electrons in a low-pressure discharge will be accelerated to high energies without making a collision. Although there was some evidence that a small number of x-rays was produced, it was not until about thirteen years later (Gibson 1957) that the electron accelerating properties of a toroidal discharge were confirmed. (These accelerated electrons are referred to as 'runaway' electrons; see Harrison 1958 and Dreicer 1959.)

It was after the end of the second world war and the successful explosion of the atom bomb that several independent workers gave serious thought to the possibility of producing controlled thermonuclear reactions. To generate useful power from such reactions heavy hydrogen must be heated to temperatures of the order of 10^8 °K. If the gas contains n positive ions per cm^3, the time t in seconds for which it must be maintained at this temperature must be such that nt is greater than 10^{16} cm^{-3} sec for deuterium and 10^{14} for a deuterium–tritium mixture (Lawson 1957). Since the temperatures required are more than 10^4 times higher than the highest temperatures which any material container could withstand without vaporizing, such a hot gas can be confined only by an electromagnetic field. This confinement is possible in principle since even well below this temperature hydrogen will be completely ionized into electrons and positive ions.

An electromagnetic field is required which will cause the gas in contact with the wall to have either a much lower density or a much lower temperature than the gas in the centre of the vessel, so that the particle bombardment of the walls involves only a small amount of heat. That is, there must be a large pressure gradient between the walls and the main body of the hot gas. From Eqn (1) it can be seen that the electromagnetic field of a high current discharge produces such a pressure gradient in the radial direction. There is, however, no pressure

gradient parallel to the current and hence no containment in this direction. If the discharge is induced in a torus the need for containment in this direction is avoided.

Sir George Thomson, then at Imperial College, was among the first to realize the potentialities of the toroidal discharge (Ware 1957, Thonemann 1958), and work started under his direction in 1947 on a discharge similar to Steenbeck's 'Wirbelrohr' (Ware 1951). Short current pulses of 2×10^4 A were achieved, and with the aid of high-speed photography, Cousins and Ware (1951) demonstrated that a marked pinch effect was occurring. Because of the speed with which the high currents were generated (in a few microseconds) the constriction of the discharge was very rapid and the inertia of the collapsing gas led to a radial oscillation of the discharge about its equilibrium configuration. This behaviour is shown in Fig. 1 (Plate I) which is a more recent streak photograph. Fig. 2 (Plate I), which is a photograph taken with a glass toroidal discharge tube, illustrates the high degree of constriction which is momentarily achieved in such discharges. Other experimental work on the pinch effect was commenced by Thonemann at Oxford in 1948 (Thonemann and Cowhig 1951) and in the U.S.A. by Tuck at Los Alamos in 1952 and Baker in Berkeley in 1952 (see Bishop 1958) and in the U.S.S.R. by Artsimovich in Moscow about 1952 (see Kurchatov 1956).

At Princeton, U.S.A., a study was commenced in 1951 by Spitzer of a high current toroidal discharge with a large magnetic field applied parallel to the current so that little pinch effect occurred (see Bishop 1958). Some early experimental work was carried out also by Bostick, Levine and Combes (1953). Subsequently many other laboratories in these and other countries have taken up the study of high current discharges. (See *The Proceedings of the Second International Conference on Peaceful Uses of Atomic Energy*, 1958, vols 31 and 32.) The work in the United Kingdom was made secret in 1950 and in the following year the research team at Oxford moved to the Atomic Energy Research Establishment at Harwell and the workers at Imperial College moved to the Associated Electrical Industries Research Laboratory, Aldermaston. The research in both the U.S.A. and U.S.S.R. was classified from its inception. As a result, despite the active research in progress, there were very few publications on the subject until 1958, when the major powers declassified their work in this field.

In order to demonstrate just a small thermonuclear reaction a temperature of a few million degrees is required and must be maintained for a few microseconds or more, depending on the particle density. For a typical number of particles per unit discharge length of 10^{17} cm^{-1}, Eqn (4) shows that the current amplitude required is about 10^4 e.m.u. or 10^5 A. Such current pulses can be produced with quite modest condenser banks storing an energy of about 2×10^4 joules. The early work in this field was therefore aimed at producing such currents and inducing a small number of reactions.

Because of the short duration of the current pulse some of the experimental studies of such discharges can be conducted in straight tubes with electrodes. In the first few microseconds the heat conducted to the electrodes will be small and contamination by electrode vapours negligible. The simpler geometry of the cylinder compared with a torus greatly facilitates experimental work as well as the theory and analysis of the results.

Instability.

The most important discovery made during the period of secrecy was that the magnetically constricted discharge is unstable. Almost any perturbation to the cylindrical symmetry or longitudinal uniformity of the discharge will cause the forces $\mathbf{j} \wedge \mathbf{B}$ and ∇p to change so that the resultant forces make the perturbation grow. Thus, for example, if the discharge remains symmetrical about the axis but becomes slightly more constricted at one or more points along its length, B_θ and j_z at these places both increase and the increase in the force $j_z B_\theta$ causes the discharge to constrict still further. The excess plasma pressure causes an expansion of the discharge elsewhere. Again, a kink in the discharge leads to an enhancement of $\mathbf{B}$ on the concave side of the kink and a reduction on the convex side, and the magnetic forces enlarge the kink.

The various geometrical configurations which these magnetohydrodynamic instabilities can take are best described by Fourier components whose form, including their time variation, is the real part of

$$A = A'(r) \exp i(\omega t - m\theta - kz) \qquad \ldots\ldots(4)$$

where A represents the change in any of the discharge parameters such as density, pressure, magnetic field, etc., $A'(r)$ is a complex number and a function of radius, m is the number of wavelengths in the θ direction in the range 0 to 2π and $2\pi/k$ the wavelength parallel to the z axis. (Complex numbers are used as a convenient means of including the phase of the particular component.) The instabilities observed experimentally have been found in most cases to correspond to a single value of m, either 0 or 1, and to have a well-defined fundamental for k. The $m = 0$ mode is often referred to as the 'sausage' instability, and the $m = 1$ mode as the 'wriggle' or 'kink' instability. Examples of these two modes are shown in Figs 3 and 4 (Plate II), Fig. 3 being a photograph with 0·2 μsec exposure and Fig. 4 is a streak photograph. Fig. 5 (Plate III) shows the corresponding instabilities induced in a falling column of mercury by passing through it a current of 300 A (Dattner, Lehnert and Lundquist 1958). The ideal discharge configuration shown in Fig. 1 persists for only a microsecond or two after initiation before it degenerates to something like that shown in Fig. 4.

In the U.K. high currents approaching those needed for a small thermonuclear reaction were achieved as early as 1950 (see Allibone *et al.* 1958), but the hydrogen discharges were found to be highly polluted with impurities due to vaporization of the tube walls. The mechanism by which the heat generated in the gas was getting to the walls was almost certainly the discharge instability, but this was not recognized at the time. It was not until 1953 that the discharge instability was discovered by Carruthers and Davenport (1957) who were working on much lower current pulses in argon. With this heavier gas and lower current a fairly regular $m = 1$ instability was observed. These instabilities were discovered independently in Russia; the date is not given in the literature but was probably 1952, since means of stabilizing the discharge were studied theoretically in 1953 (Braginsky and Shafranov 1958). A further independent discovery was in the work of Granovskii and Timofeeva (1955, 1956). In the U.S.A. it was predicted theoretically in the work of Kruskal and Schwarzschild (1954) before it was observed experimentally—a notable achievement.

The magnetohydrodynamic instabilities bring the plasma into contact with the tube wall and lead to extensive particle and heat losses and to wall vaporization. They have been the major obstacle to the achievement of high temperatures and probably still are. Many attempts were made to inhibit the instability with various types of applied magnetic field. In the U.K., where fairly low current discharges (10^3 to 10^4 A) with low rates of rise of current were used for instability studies, a simple magnetic field applied parallel to the discharge increases instability (Allibone *et al.* 1958). However, Bickerton (1957), working with higher currents (5×10^4 A) and a metal torus, found that a longitudinal field caused a marked decrease in the amplitude of the instability fluctuations. It was shortly afterwards that the theoretical work of Tayler (1957 b) showed that a discharge will be stable if a longitudinal magnetic field exists within the constricted discharge with little or no longitudinal field outside, and if it is surrounded at not too great a distance by a metal wall. Corresponding developments occurred in the U.S.A. and U.S.S.R., culminating in the similar theoretical papers of Rosenbluth (1956) and Shafranov (1956). The equations of motion for an ionized gas in a magnetic field are very complex, and these theoretical advances were possible only when a very simplified theoretical model was taken. The model was a cylinder of perfectly conducting fluid with all currents flowing on the surface of the cylinder.

Discharges with configurations corresponding roughly to this theoretical model were produced in straight tubes (Bezbatchenko *et al.* 1956, Burkhardt, Lovberg and Phillips 1958, Hagerman and Mather 1958, Colgate, Ferguson and Furth 1958), and in the toroidal discharge experiments ZETA (Thonemann *et al.* 1958), SCEPTRE III (Allen *et al.* 1958) and PERHAPSATRON S3 (Honsaker *et al.* 1958). In the case of the discharges with rapid rise of current (rise times of a few microseconds), such as studied in the straight tubes and PERHAPSATRON S3, a distinct onset time is observed for instability. The stabilizing magnetic field was found to increase the onset time and to reduce the instability to a much smaller amplitude when it did occur. With the slow discharges no sharp onset is observed, but the stabilizing field caused a marked reduction in the instability amplitude. The substantial advance in stability in these experiments was accompanied by marked improvements in other discharge parameters including the electron and ion temperatures, impurity level and the nuclear reaction rate, particularly in the case of the toroidal experiments. Because of both the lower level of impurity and the higher electron temperature the visible radiation from these discharges was reduced and normal high-speed photography ceased to be able to detect the discharge channel when hydrogen was used. Time-resolved spectroscopy (Allibone *et al.* 1958, Butt *et al.* 1958) revealed that there is a high-temperature core to the discharge which moves very little and exhibits only small fluctuations (see Fig. 6). However, normal high-speed photography has shown that outside this channel there is a considerable amount of colder plasma which emits light with fluctuating intensity (Allibone *et al.* 1958, Butt *et al.* 1958). On rotating-mirror streak photographs this light is sometimes seen to follow helical patterns and on other occasions bursts of light appear simultaneously at all points across the viewing slit, so that 'bars' are seen on the streak photograph, as shown in Fig. 7 (Plate III). There is a close correlation between these light fluctuations and the magnetic field fluctuations (Ware 1959 b).

Simultaneously with this experimental work, the theoretical physicists were improving the simple model used in their analysis and were allowing for the fact that in practice the discharge currents are distributed over a sheath with finite radial thickness. Their results showed that unless the skin is very thin (Boon *et al.* 1958, Hubbard 1958) or the magnetic field varies with radius according to fairly stringent conditions (Rosenbluth 1958, Suydam 1958), the discharge will be unstable within the current sheet. The magnetic field configurations in all

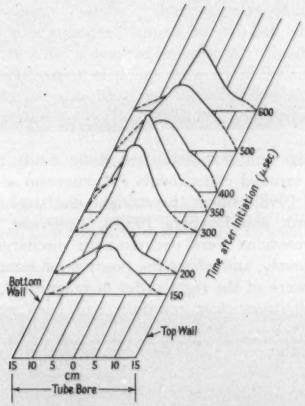


Fig. 6. Spatial distribution of spectral line intensity from four times ionized oxygen at different times throughout the current pulse with SCEPTRE III (Allibone *et al.* 1958).

the above experiments did not satisfy these new conditions; hence these layer instabilities should occur and they present the simplest explanation of the small amplitude fluctuations which are still observed. More recent experimental results favour this assumption (Birdsall *et al.* 1959 a, Tuck 1959, Ware 1959 b, c).

Nuclear reactions.

Deuterium fusion reactions were detected in 1952 in the linear discharges studied in the U.S.S.R. (Kurchatov 1956) and in 1955 in similar discharges in the U.S.A. (Anderson *et al.* 1957). In these experiments with no axial magnetic field, current pulses with amplitude about 10^5 A were generated in a few microseconds so that very rapid constriction to a small diameter occurred, followed by several oscillations about the equilibrium diameter. The nuclear reactions—detected by the neutrons emitted—occurred at the second and sometimes subsequent times of maximum constriction. Since the rapid collapse generates a shock wave, considerable heating of the positive ions was expected and there were hopes that the fusion reactions were thermonuclear, that is, induced by a high temperature. Subsequently the energy of neutrons emitted parallel to the current

was found to be slightly in excess of the energy of those emitted in the opposite direction. This non-isotropy indicated a centre-of-mass velocity for the reacting deuterons of about 10^8 cm sec^{-1} parallel to the current. This result is most easily explained by assuming that a small proportion of the deuterons in the discharge are accelerated by some extraneous mechanism to high energies, and that some of these particles react with relatively stationary deuterons. The most likely cause of the particle acceleration is the $m = 0$ instability (Anderson *et al.* 1957), which causes large inductive voltages at the points of maximum constriction.

The number of neutrons generated in these experiments was found to be greatly reduced by the addition of a small percentage of an impurity gas or a small axial magnetic field. In other experiments with very low external circuit inductance, the nuclear reactions were much less sensitive to both impurity and axial magnetic field (Allen, K. W., *et al.* 1957, Curran *et al.* 1958). However, neutron energy measurements again showed a moving centre of mass for the reacting particles.

In the experiments with stabilizing magnetic fields neutron emission was observed in both the toroidal experiments (Thonemann *et al.* 1958, Allen *et al.* 1958, Honsaker *et al.* 1958) and in the straight discharge tubes (Hagerman and Mather 1958, Burkhardt and Lovberg 1958 a, Anderson, Baker, Colgate *et al.* 1958). Since these reactions were occurring in discharges which were much more stable than previously, and where the positive ion temperature from spectroscopic measurements were of the right order of magnitude, there was again considerable hope that thermonuclear reactions had been achieved. Once again, however, the neutron emission was found to be anisotropic. The cause of the particle acceleration associated with these reactions has yet to be satisfactorily explained.

One of the other types of high current discharge studied in the past few years has been a cylindrical discharge in which a very rapidly rising axial magnetic field induces a high current in the θ direction (Post 1958, Colgate and Wright 1958, Elmore *et al.* 1958, Kolb 1958, Osovets *et al.* 1958). The field is often increased in magnitude at each end of the cylinder to provide magnetic mirrors to reduce end losses. The detailed mechanism of this type of discharge (see § 8) is somewhat different from the pinch discharge but an analogous constriction of the plasma occurs. At the time of The Second International Conference on the Peaceful Uses of Atomic Energy (Geneva, September 1958) this was the only discharge for which a possible thermonuclear reaction was claimed (Elmore *et al.* 1958). Since then these discharges have been studied in considerable detail, and although the heating of the deuterons may well be due to an instability (Kolb, Dobbie and Griem 1959), there is now very strong evidence that the reactions are truly thermonuclear (Ribe 1959). These discharges are also unique for another reason : they are the only discharges in which electrons have been heated to sufficiently high temperatures with sufficiently high densities for their bremsstrahlung radiation to have a detectable intensity at soft x-ray wavelengths† (Griem, Kolb and Faust 1959, Boyer *et al.* 1959). (Bremsstrahlung photons are those generated by free–free collisions between electrons and positive ions.) It is not

† This radiation is now believed to be the recombination spectra of highly ionized impurity ions (Jahoda *et al.* 1960).

PLATE I

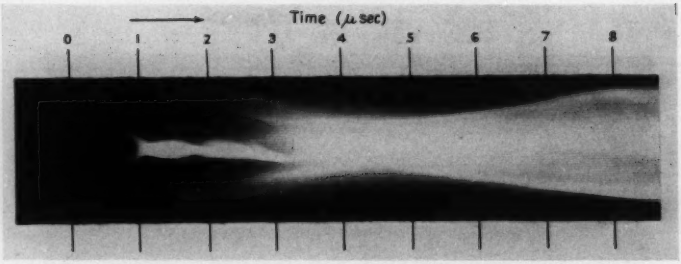


Fig. 1. Streak photograph of a deuterium pinch discharge with 5% oxygen impurity taken by Gabriel (1959), showing the rapid contraction and ' bouncing ' of the discharge. The formation of a second current sheet can also be seen starting at 1·5 μsec.

Fig. 2. Toroidal pinch discharge in krypton. This photograph was taken by Burkhardt *et al.* (1957) using an open shutter. It shows the high degree of constriction momentarily achieved, at which time the discharge emits much more light.

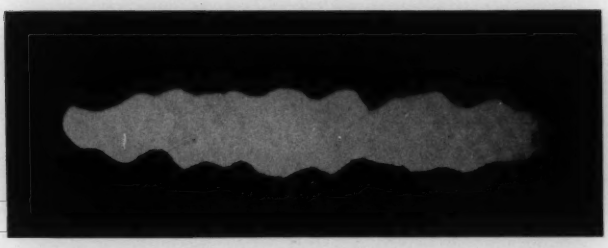


Fig. 3. Example of an $m = 0$ or ' sausage ' instability in a linear pinch discharge in argon. This photograph was taken by Latham, Cuzon and Nation (1959) using an image converter with 0·2 μsec exposure time.

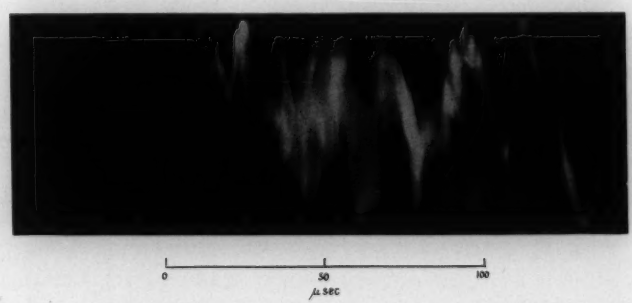


Fig. 4. Streak photograph of an $m = 1$ or ' wriggle ' instability. This violent lashing of the discharge occurs at a later time than Figs 1, 2 and 3. Taken from Ware (1957).

PLATE III

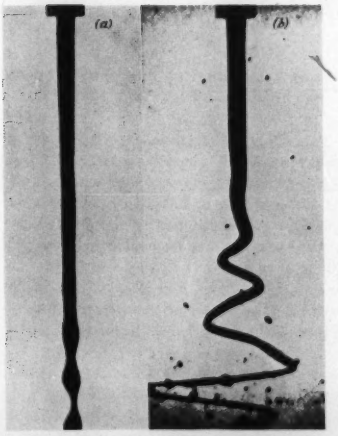


Fig. 5 (*a*) and (*b*). Examples of ' sausage ' and ' wriggle ' instabilities in a column of mercury. Because the mercury is falling, the instability is further developed at the bottom (Dattner, Lehnert and Lunquist 1958). Part (*a*) was taken with no axial magnetic field and (*b*) with an axial magnetic field of 300 gauss.

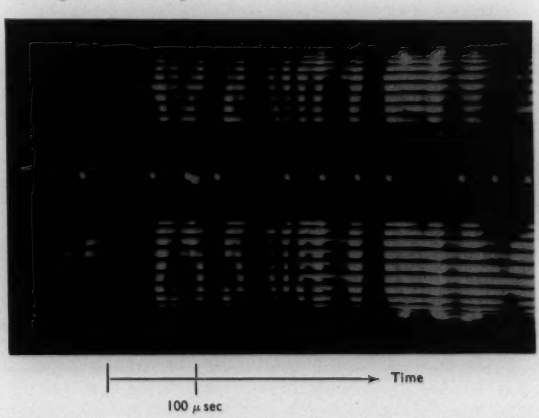


Fig. 7. Streak photographs of a hydrogen discharge in SCEPTRE III, showing the so-called ' bars '. (The two streak pictures are different views of the same discharge.) Taken from Payne and Kaufman (1959).

surprising that the study of this type of discharge has been taken up by many other laboratories. (See *The Proceedings of the Fourth International Conference on Ionization Phenomena in Gases*, Uppsala, 1959.)

Energy loss.

A characteristic feature of the high current discharge is the very large rate at which energy is generated in the discharge due to ohmic heating. Instantaneous powers as high as 900 MW have been quoted, and since this power is being dissipated in an amount of gas weighing less than 10^{-3} g, there is clearly no problem at present in delivering heat to the discharge. In many experiments the amount of energy introduced by the time peak current has been reached has been enough to heat all the particles in the gas to several tens of millions of degrees. Even in the presence of a stabilizing magnetic field and in the toroidal discharge experiments where there are no end losses, the amount of thermal energy retained in the gas is at least an order of magnitude less than this figure (Colgate, Ferguson and Furth 1958, Conner *et al.* 1958, Allen and Liley 1959). Thus, energy is being lost from the discharge at roughly the same rate as it is delivered. The search for the mechanism of this heat loss has been the main problem associated with these discharges since early in 1958.

The first workers to study systematically the possible heat-loss processes were Colgate, Ferguson and Furth (1958). In their 'gamma' toroidal pinch experiment, during the first eighth of a cycle (about 5 μsec) the main heat loss was found to be due to energetic electrons (energy $\sim 1{\cdot}5$ kev) striking the tube walls. After the first quarter of a cycle the amount of ultra-violet radiation due to impurity ions in the discharge became large and was the main energy loss mechanism throughout the rest of the current pulse. It was inferred that electrons with lower energies were being lost during the second eighth of a cycle, since the measuring technique could not detect electrons with energy less than 1 kev.

These results suggest that the fundamental problem is electron loss from the discharge. Since the electron bombardment will cause the quartz discharge tube used in this experiment to vaporize, the impurity radiation later in the pulse is probably a secondary effect. More recent results of other workers are in general agreement with these early conclusions (see § 6.1). It has been proposed by Ware (1959c) that the hydromagnetic instability provides the mechanism by which the electrons traverse the confining magnetic fields.

§ 2. The Plasma Equations and some Physical Concepts

2.1. *A Set of Plasma Equations*

The mechanics of an ionized gas, even when it is completely ionized hydrogen, is a very complicated subject. This is because such a gas is a mixture of two types of particles (electrons and positive ions) and because it has many characteristic frequencies, such as the electron and positive cyclotron frequencies, the electron plasma frequency and the electron–electron, electron–ion and ion–ion collision frequencies. Last, and not least, is the fact that the equations of motion are non-linear. The only comprehensive and exact equations are the Boltzmann equations for the velocity distribution functions of the electrons and positive ions.

If $f_i(\mathbf{w}, \mathbf{r}, t)\, dw_x\, dw_y\, dw_z\, dxdydz$ is the number of ions in the volume element $dxdydz$ at time t whose velocity components w_x, w_y, w_z are in the ranges dw_x, dw_y, dw_z, the Boltzmann equation for f_i is

$$\frac{\partial f_i}{\partial t} = -\mathbf{w}.\nabla f_i - \frac{e}{M}\left(\mathbf{E} + \frac{1}{c}\mathbf{w} \wedge \mathbf{B}\right).\nabla_w f_i + \left(\frac{\partial f_i}{\partial t}\right)_{\text{coll}} \qquad \ldots\ldots(5)$$

(see Chapman and Cowling 1939). The first term on the right-hand side is the rate at which ions leave the volume element $dxdydz$ due to their velocity, the second term is the rate at which ions leave the corresponding volume element $dw_x\, dw_y\, dw_z$ in velocity space due to their acceleration $e(\mathbf{E} + \mathbf{w} \wedge \mathbf{B}/c)/M$ and the third term is the change in f_i due to particle collisions. There is a corresponding equation for the electron velocity distribution function f_e.

In order to solve these equations and obtain transport equations relating such macroscopic parameters as pressure, mass velocity and acceleration, current density, etc., it is necessary to make several simplifying assumptions, and even then the solutions are obtained only by a series of successive approximations. The order of approximation achieved has generally been dictated by the mathematical difficulty of proceeding further. Since a particular set of approximations will not have general validity, the transport equations obtained will apply for only a limited range of discharge conditions.

Here a set of transport equations are presented which are obtained from the Boltzmann equations assuming that the distribution functions are only slightly perturbed from Maxwellian distributions. In the case of phenomena which occur at frequencies low compared with the particle collision frequencies the equations will be a good first-order approximation. In particular they will probably be valid for most of the hydromagnetic instabilities observed so far, since instability frequencies have generally been less than the collision frequencies. The plasma equations have been taken from Chapman and Cowling (1939) and Liley (1959), but they are similar to those which can be obtained from Landshoff (1949) or Spitzer (1956). The equations are for completely ionized hydrogen or deuterium. The reader is referred to the above references for the changes which must be made when a gas with higher ionic charge is considered. For completeness the continuity equation and Maxwell's equations for the electromagnetic fields are given. Since we are dealing with low frequencies the displacement current has been ignored. Lastly, radiation has been neglected.

Plasma momentum balance :

$$\rho \frac{d\mathbf{V}}{dt} = \frac{1}{c}\mathbf{j} \wedge \mathbf{B} - \nabla p \qquad \ldots\ldots(6)$$

where

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{V}.\nabla.$$

Electric conduction equation :

$$\mathbf{E} + \frac{1}{c}\mathbf{V} \wedge \mathbf{B} - \frac{1}{nec}\mathbf{j} \wedge \mathbf{B} + \frac{1}{ne}\nabla p_e = \eta \mathbf{j} - \alpha \mathbf{q}_e. \qquad \ldots\ldots(7)$$

Electron heat flux equation :

$$\frac{5p_e k}{2m}\nabla T_e + \frac{e}{mc}\mathbf{q}_e \wedge \mathbf{B} = -\beta\mathbf{q}_e - \gamma\mathbf{j}. \qquad \ldots\ldots(8)$$

Thermal energy equations :

$$\frac{\partial T_e}{\partial t} + \mathbf{v}_e.\nabla T_e + \frac{2}{3kn}\{p_e\nabla.\mathbf{v}_e + \nabla.\mathbf{q}_e\} = \frac{2}{3kn}\mathbf{j}.\left(\mathbf{E} + \frac{1}{c}\mathbf{V}\wedge\mathbf{B}\right) + \frac{1}{\tau_{ei}}(T_i - T_e) \qquad \ldots\ldots(9)$$

$$\frac{\partial T_i}{\partial t} + \mathbf{V}_i.\nabla T_i + \frac{2}{3kn}\{p_i\nabla.\mathbf{V}_i + \nabla.\mathbf{q}_i\} = \frac{1}{\tau_{ei}}(T_e - T_i). \qquad \ldots\ldots(10)$$

Continuity equation :

$$\frac{\partial n}{\partial t} = \nabla.(n\mathbf{V}_i) = \nabla.(n\mathbf{v}_e). \qquad \ldots\ldots(11)$$

Maxwell's equations :

$$\nabla\wedge\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t}, \qquad \ldots\ldots(12)$$

$$\nabla\wedge\mathbf{B} = 4\pi\mathbf{j}/c, \qquad \ldots\ldots(13)$$

$$\nabla.\mathbf{B} = 0. \qquad \ldots\ldots(14)$$

Gaussian units have been used and $\rho = n(M+m)$ is the plasma density, M the ion mass, m the electron mass, n the number of electrons per unit volume approximately equal to the number of positive ions per unit volume, the mass velocity $\mathbf{V} = (M\mathbf{V}_i + m\mathbf{v}_e)/(M+m)$, $\mathbf{V}_i$ the mean ion velocity, $\mathbf{v}_e$ the mean electron velocity, $\mathbf{j}$ the current density, $p = p_i + p_e$ is the total plasma pressure, p_i the positive ion pressure, p_e the electron pressure, $\mathbf{E}$ the electric field, $\mathbf{B}$ the magnetic field, c the velocity of light, e the proton charge, η the resistivity perpendicular to the magnetic field when no Hall current flows $= 1{\cdot}43\times10^{-9}\ln\Lambda/T_e^{3/2}$ e.s.u. $(= 1{\cdot}29\times10^{4}\ln\Lambda/T_e^{3/2}\ \Omega cm)$, $\Lambda = (3/2e^3)(k^3T_e^3/\pi n)^{1/2}$, $\alpha = 3\eta e/5kT_e$, $\mathbf{q}_e$ the heat flux due to electrons, referred to the mean electron velocity, $\mathbf{q}_i$ the heat flux due to positive ions, referred to the mean ion velocity, k Boltzmann's constant, T_e the electron temperature, T_i the ion temperature, $\beta = 1{\cdot}9\eta ne^2/m$, $\gamma = 1{\cdot}5\eta enkT_e/m$, and $\tau_{ei} = [3mMk^{3/2}/8(2\pi)^{1/2}ne^4\ln\Lambda](T_e/m + T_i/M)^{3/2}$.

Equation (6) is the well-known equation of motion for the plasma and shows that the mass acceleration is given by the difference in the two forces on the right-hand side. The electric conduction equation (sometimes called the generalized Ohm's law) is essentially the equation of motion for the electrons with the acceleration term neglected because of the small electron mass. The first three terms constitute the effective electric field acting on the moving electrons ; the second term is the contribution due to the mean plasma velocity and the third term (the Hall e.m.f.) allows for the difference between the mean electron velocity and the plasma velocity. The fourth term is the e.m.f. due to the electron pressure gradient. The terms on the right-hand side are proportional to the mean force acting on the electrons due to collisions with the positive ions, one term being proportional to the electron current and one to the electron heat flux.

In Eqn (9) the first two terms constitute the total rate of change with time of the temperature of a particular group of electrons. The first term in the brackets on the left-hand side gives the change of T_e due to compression or expansion of the electron gas, and the second term in the brackets is the change due to the divergence of the electron heat flux. The two terms on the right-hand side represent respectively the heating of the electrons due to the ohmic heating of the current and due to energy transfer from the positive ions by collisions. Eqn (10) is the corresponding equation for T_i, except that the ohmic heating of the ions is small and has been neglected.

In some cases, such as when a large mass–velocity gradient exists, the isotropic pressure p must be replaced by a pressure tensor so as to include the effect of viscosity. In other cases, such as in discharge models which assume a plasma edge, it will be necessary to allow for a net volume or surface electric charge in the plasma. The appropriate terms in the plasma equations are given by Liley (1959). The fourth Maxwell equation (for $\nabla . \mathbf{E}$) will also be required.

It should be noted that the electron heat-flux equation does not include the positive ion contribution to heat conduction. In the absence of a magnetic field the contribution to the thermal conductivity due to protons in ionized hydrogen is about 1/15 of the electron component (Chapman 1954). Since, however, in the presence of a magnetic field the electron heat conduction is reduced by the factor $1/(1+\omega_e^2\tau_e^2)$ where ω_e is the electron cyclotron frequency and τ_e the mean electron collision frequency, quite modest magnetic fields will reduce the electron component below that for the ions. The positive ion heat-flux equation is given by Braginsky (1957 a).

2.2. *The Magnetohydrodynamic Approximations*

Although many simplifying assumptions have already been made in deriving the above equations, their solutions for a particular problem would still be a very formidable mathematical task. Progress has been made possible only by making further assumptions. Most notable of these assumptions are: (*a*) the neglect of the Hall e.m.f. ($\mathbf{j} \wedge \mathbf{B}/ne$) and the electron pressure gradient e.m.f. ($\nabla p_e/ne$) from the electric conduction equation, and (*b*) the neglect of heat conduction from the thermal energy equation.

With these further approximations the set of equations becomes independent of the fine structure of the plasma and they apply equally to a continuous conducting fluid; they are known as the magnetohydrodynamic equations.

A further approximation which is frequently made is to neglect the two terms in the electric conduction equation which are proportional to the collision frequency, namely $\eta\mathbf{j}$ and $\alpha\mathbf{q}_e$. Together with the approximations (*a*) above, Eqn (7) reduces to the simple form associated with a perfect metallic conductor, namely

$$\mathbf{E} + \frac{1}{c}\mathbf{V} \wedge \mathbf{B} = 0. \qquad \ldots\ldots(15)$$

2.3. *The Collisionless Plasma Approximation*

For effects which occur rapidly compared with the particle collision frequency, the assumption of approximately Maxwellian velocity distributions will not be

valid. This will often be the case as higher temperatures are achieved since the collision frequency is proportional to $T^{-3/2}$. When the motions of the plasma are rapid compared with the collision frequency there will be little coupling between the particle velocities parallel to the magnetic field and those perpendicular to the field. The distribution function will cease to be isotropic and the component of pressure parallel to the field $p_\parallel$ may be appreciably different from the perpendicular components $p_\perp$. The problem of solving Boltzmann's equations for these conditions has been tackled by several workers (Chew, Goldberger and Low 1956, Watson 1956, Chandrasekhar, Kaufman and Watson 1957, Rosenbluth and Rostoker 1958, Kruskal and Oberman 1958, Sagdeyev *et al.* 1958). Although the procedures differ in other respects, they have all assumed that the magnetic field is large and obtained solutions for the transport equations by using a series expansion in terms of the parameter ρ_L/L or ω/ω_i, where ρ_L is the particle Larmor radius, L is the characteristic length over which quantities change in the discharge, $\omega/2\pi$ is the frequency with which quantities change and $\omega_i/2\pi$ is the ion cyclotron frequency. The transport equations obtained in this way have been used mainly in the theory of discharge instability. At the degree of approximation to which the solutions have been taken the results obtained have not been greatly different from those obtained with the magnetohydrodynamic equations. One reason for the similarity in the results is that to first order in the expansion parameter the electric conduction equation is identical with the perfect conductivity approximation (Eqn (15)).

2.4. *Simple Physical Concepts*

In order to assist the theoretical physicist in choosing the most suitable approximations, and to assist the experimenter in understanding his results it has been necessary to conceive simple models to replace the complex plasma. The perfectly conducting compressible cylinder is an example. In order to visualize the physical processes involved, several physical concepts have been of great assistance. Most notable among these are the concepts of 'magnetic pressure' and the 'freezing in' of magnetic lines of force; these are summarized below. For fuller descriptions of these and other plasma processes the reader is referred to the works of Spitzer (1956) and Cowling (1957).

2.4.1. *Magnetic pressure.*

(*a*) *Magnetic fields with no curvature.* When the magnetic lines of force are straight it follows directly from Eqn (13) that the force $\mathbf{j} \wedge \mathbf{B}$ is equal to $-\nabla(B^2/8\pi)$. In this case therefore the quantity $B^2/8\pi$ acts as a scalar pressure, and for equilibrium, Eqn (5) can be integrated to give

$$p + \frac{B^2}{8\pi} = \text{const.} \qquad \ldots\ldots(16)$$

(*b*) *Magnetic fields with curvature.* In this case the concept of a simple scalar pressure breaks down. Taking as an example the pinch effect where there is cylindrical symmetry and the only component of $\mathbf{j}$ is j_z, the force is

$$-j_z B_\theta = -\frac{\partial}{\partial r}\left(\frac{B_\theta^2}{8\pi}\right) - \frac{B_\theta^2}{4\pi r}. \qquad \ldots\ldots(17)$$

The extra term arises because magnetic tubes of force behave as if they have a tension along their length of $B^2/8\pi$ per unit area in addition to the pressure in the lateral direction. In the special case where a pinch discharge has a current sheath thin compared with the radius, Eqn (17) can be integrated over the sheath thickness δ giving

$$\int_r^{r+\delta} -j_z B_\theta \, dr \simeq -\frac{B_{\theta e}^2}{8\pi} \qquad \ldots\ldots(18)$$

where $B_{\theta e}$ is the field just outside the sheath and where the field inside the sheath is assumed zero. In this special case the magnetic field can once again be regarded as a pressure $B_{\theta e}^2/8\pi$ pushing inwards on the discharge.

2.4.2. *'Freezing in' of magnetic lines of force.*

When the perfect conductivity approximation is made, a simple relationship can be obtained by taking the curl of Eqn (15), and substituting for $\nabla \wedge \mathbf{E}$ and $\nabla . \mathbf{V}$ from Eqns (12) and (11) (Cowling 1957, page 7), namely,

$$\frac{d}{dt}\left(\frac{\mathbf{B}}{n}\right) = \left(\frac{\mathbf{B}}{n}\right).\nabla\mathbf{V}. \qquad \ldots\ldots(19)$$

Thus for motion of the plasma for which the component of $\nabla\mathbf{V}$ parallel to $\mathbf{B}$ is zero, $\mathbf{B}/n$ is a constant of the mass motion, and the lines of force can be regarded as frozen to the plasma and moving with the mass velocity.

If, however, the more exact conduction equation (7) is used, a substantially different result is obtained. Assuming that the collision frequency is low enough to make the two terms on the right-hand side of (7) negligible, the relation corresponding to Eqn (19) is (Schluter 1956)

$$\left(\frac{d}{dt}\right)_e \left(\frac{\mathbf{B}}{n}\right) = \frac{\mathbf{B}}{n}.\nabla\mathbf{v}_e - \frac{1}{n^2 e}\nabla n \wedge \nabla p_e \qquad \ldots\ldots(20)$$

where

$$\left(\frac{d}{dt}\right)_e = \frac{\partial}{\partial t} + \mathbf{v}_e.\nabla.$$

In this case, if $\nabla n \wedge \nabla p_e$ is small, for motion such that the component of $\nabla\mathbf{v}_e$ parallel to $\mathbf{B}$ is zero, the magnetic field is frozen to the electrons and the lines of force move with the mean electron velocity. If $\nabla n \wedge \nabla p_e$ is not zero, there is a generation of magnetic field within the plasma.

When the finite conductivity of the plasma is taken into account the lines of force are found to diffuse out of a region in which they are frozen. The time constant for the exponential decay of the field from a region of dimension L is $4\pi L^2 c^2/\eta$ (Spitzer 1956, page 38).

§ 3. Discharge Initiation and Contraction

3.1. *Ionization*

The initial formation of high current discharges is particularly complex due to the many processes taking place simultaneously with different time constants. The main processes can be listed as follows :

(i) ionization,

(ii) radiation,

(iii) heating of (*a*) the electrons and (*b*) the ions,
(iv) increase in current,
(v) penetration of the magnetic field (or current) into the discharge,
(vi) contraction of discharge (and formation of a shock wave).

Because of the number and complexity of these processes and the coupling between them it is not surprising that a comprehensive theory has not been undertaken. In certain cases it has been possible to neglect several of the processes, and even then a substantial problem for an electronic computer remains.

With a few exceptions, very little work has been done on the ionization processes. One notable exception has been the computation of the ionization and heating of a very low electric field (0·1 v cm^{-1}) and low pressure (10^{-4} mm Hg) discharge in helium (Berger *et al.* 1958). The electric field was assumed constant across the discharge and with respect to time, and contraction of the discharge was assumed inhibited by a large axial magnetic field, so that processes (v) and (vi) were eliminated. The currents, temperatures and spectral line intensities were obtained throughout the first and second stages of ionization. Kogan (1958) has developed a method of estimating the electron temperature during the ionization and contraction stage of the linear discharge with high rate of rise of current; the electron temperature during this phase is kept down by the ionization process and the values found were in the range 2·5 to 4 ev, depending on the initial hydrogen pressure (Borzunov, Kogan and Orlinsky 1958). Corresponding calculations for the slow toroidal pinch discharge have been made by McWhirter (1959) and measurements by McWhirter, Griffin and Jones (1959).

Interest has been shown recently in the ionization process in connection with choosing the optimum electric field for most rapid ionization. That the electric fields can be too high as well as too low for rapid ionization is shown clearly in the breakdown curves obtained by Etievant *et al.* (1959) for a toroidal discharge. The growth of ionization at high electric fields has been studied theoretically by Gerjuoy and Stuart (1960).

3.2. *Discharge Contraction*

3.2.1. *High rate of rise of current, no axial magnetic field.*

High current discharges have been studied with rates of rise of current (dI/dt) varying over the wide range 10^7 to 10^{12} A sec^{-1}. It is not surprising that different initial behaviour has been observed. In the case of high rates of rise of current, where the high applied electric field causes a high degree of ionization in a time short compared with the discharge contraction time, and where the rapid rise of current limits the current-carrying region to a thin annular sheath on the outside of the discharge (the skin effect), a fairly well-defined type of discharge contraction occurs. The high current density in the sheath rapidly ionizes most of the gas in this region and the $\mathbf{j} \wedge \mathbf{B}$ force gives this plasma an inward acceleration. Since the force operates only on the outer layer, a high pressure is built up on the inside of the contracting sheath.

The sheath will accelerate rapidly to a terminal velocity close to the value cE_z/B_θ where E_z and B_θ are the electric and magnetic fields at the edge of the discharge. This is the velocity at which the back e.m.f. IdL/dt balances the

electric field at the edge of the discharge. (L is the inductance per unit length of the discharge.) If this velocity is comparable with or exceeds the speed of sound for the gas in the central region of the discharge, the high pressure will propagate inwards as a shock wave, the shock front having a velocity somewhat higher than the current-carrying sheath (see Fig. 8). The outer current-carrying edge can be regarded as the contact front between the magnetic pressure and the gas. The shock front is 'reflected' from the axis of the discharge and then travels radially outwards. When it meets the incoming edge of the discharge it causes the edge to change its radial velocity from a negative to a positive value. The

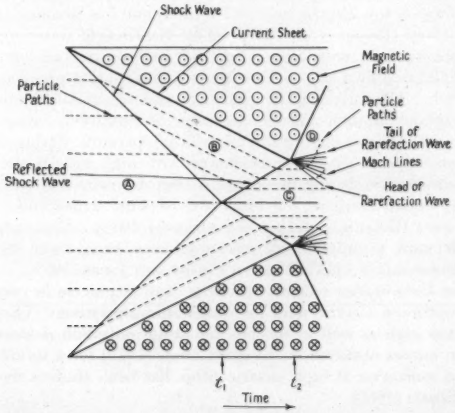


Fig. 8. Diagrammatic representation of the pinch discharge with high rate of rise of current, showing the contracting discharge edge and the shock wave preceding it. Acceleration has been ignored. Taken from J. E. Allen (1957).

discharge edge moves outwards until the $\mathbf{j} \wedge \mathbf{B}$ force dissipates the inertia of the shock wave. The discharge then contracts again and a series of these 'bounces' occurs. As many as six bounces have been observed (Siegbahn and Ohlin 1958) before instability caused the discharge to break up. In spectroscopic studies Lukyanov and Sinitsyn (1956, 1958, 1959) have measured densities up to thirty times the initial gas density and ion energies corresponding to 10^6 to 2×10^6 °K.

The shock wave nature of the contraction can be seen in streak photographs of such discharges (for example, Cousins and Ware 1951, Allen and Reynolds 1957) and has been analysed by J. E. Allen (1957), from whose paper Fig. 8 is taken, and by Jukes (1958), who allowed for the cylindrical nature of the shock wave. If the contact and shock fronts are assumed to have roughly equal velocities, a simple 'snow-plough' model can be adopted and the equation of motion for the sheath can be solved simultaneously with the current equation (Rosenbluth, Garwin and Rosenbluth 1954). If the applied electric field E_0 is assumed constant

throughout the contraction, although the equations have to be integrated numerically, a simple formula is obtained for the time to maximum constriction t_p, namely

$$t_p = b(4\pi\rho_0/c^2 E_0^2)^{1/4} \quad \ldots\ldots(21)$$

where b is the tube radius and ρ_0 the initial gas density. The results show reasonable agreement with experiment (Anderson *et al.* 1957, Sundstrom and Svennerstedt 1959). A more elaborate model analysed by Braginsky and Migdal (1958), who allowed for finite electrical conductivity and the compression and heating of the gas, showed similar reasonable agreement with experiment (Andrianov *et al.* 1958). A further theoretical treatment is given by Kulikovskii (1957), and the trapping of the neutral gas in the collapsing discharge has been studied by Phillips (1959).

In the case where the contraction of the discharge sheath is slow compared with the speed of sound for the gas in the centre of the tube, sound waves will equalize the pressure throughout the volume within the discharge sheath and the gas will compress approximately adiabatically. No shock wave will occur but there will still be bouncing due to the inertia of the contracting discharge. An analysis appropriate to this type of contraction was done by Leontovich and Osovets (1956). Since the time taken for the discharge to contract is proportional to the fourth root of the discharge mass and the square root of the current, it is not very sensitive to the type of model taken for the discharge.

The rapid contractions and expansions of the discharge cause large changes in the discharge inductance and because the external electric circuit has appreciable impedance the e.m.f. IdL/dt causes large fluctuations in the voltage across the discharge tube (Artsimovich, Andrianov, Bazilevskaia *et al.* 1956, Anderson *et al.* 1957). These fluctuations enable the times of maximum constriction to be accurately determined.

Several experiments indicate that some neutral gas is left behind by the contracting discharge or is being evaporated from the tube walls. Magnetic probe measurements have shown that although the current density is largest in the luminous constricted discharge channel, there is some current density throughout the whole tube. The current density in the outer regions is often negative (Artsimovich *et al.* 1956, Burkhardt and Lovberg 1958a) and in some cases a further positive current sheath appears on the outside of the negative current region and moves towards the centre of the tube (Artsimovich *et al.* 1956, Komelkov *et al.* 1958, Bodin, Peacock and Reynolds 1959). The existence of this second current sheath has been observed photographically with discharges in impure deuterium (Gabriel 1959, see Fig. 1) and in gases other than deuterium (Borzunov, Orlinsky and Osovets 1958, Bodin, Peacock and Reynolds 1959). It has also been observed by Latham, Cuzon and Nation (1959), who found that it no longer occurred when the quartz tube was shielded on the inside by a series of overlapping copper cylinders. This last result is strong evidence that the gas is coming from the tube wall, since copper will undergo a much smaller temperature rise than quartz for a given heat pulse (Craston *et al.* 1958). When a Pyrex glass discharge tube is used, the initial discharge stays close to the tube wall for a longer time (Anderson *et al.* 1957) and at very large dI/dt a similar effect may be occurring with quartz tubes since the contraction times are longer than predicted by theory (Curran *et al.* 1958).

The explanation for a negative current density occurring while the applied electric field is still positive may be a positive radial pressure gradient caused by gas evolved from the tube wall. When the applied electric field reverses and the total discharge current is decaying, the negative currents in the outer regions of the tube are observed to become much larger, while the positive current in the core decreases only slowly and persists well into the second half cycle (Burkhardt and Lovberg 1958a). This latter behaviour is the same skin effect phenomenon which occurs in any thick electrical conductor. It has been analysed by Haines (1959).

In several experiments (Komelkov 1958, Bodin, Peacock and Reynolds 1959, Gabriel 1959) multiple filaments have been observed in the contracting discharge sheath.

3.2.2. *High rate of rise of current with an axial magnetic field.*

The presence of an axial magnetic field has two main effects on the contraction of the discharge. Firstly, since the current sheath is a good conductor its radial motion will cause currents in the θ direction on the inner edge of the sheath so as to maintain a constant axial magnetic flux within the cylindrical sheath. The magnetic field increases as the discharge constricts so that $\pi a^2 B_z$ is a constant, where a is the radius of the sheath. This increase in B_z within the discharge causes a back e.m.f. which will either make the current in the coil producing the initial B_z fall close to zero, or induce a negative current in the metal shield if such a shield is present which is continuous in the θ direction. In either case the magnetic field outside the discharge falls to a low value. Therefore, as far as the axial magnetic field is concerned there is an excess pressure within the discharge sheath of approximately $B_z^2/8\pi$, and the discharge will not constrict and achieve this state until the pinch effect pressure $B_{\theta e}^2/8\pi$ is equal to $B_z^2/8\pi$.

Secondly, if the gas in the central part of the discharge is sufficiently ionized to be a reasonably good electrical conductor, the presence of the magnetic field increases the speed of sound in this region to $(C_s^2+B_z^2/8\pi\rho)^{1/2}$ where C_s is the speed of sound in the absence of B_z. In addition, the rate of contraction is much slower because the pinch force $j_z B_\theta$ is still counteracted to a large extent by the outward force $j_\theta B_z$. As a result, in most experiments the contraction velocity has been small compared with the sound speed and the contraction has been of the adiabatic type referred to in § 3.2.1 above with no shock wave.

Detailed studies using magnetic probes have been made of this type of discharge contraction with straight discharge tubes (Burkhardt, Lovberg and Phillips 1958, Burkhardt and Lovberg 1958b, Colgate, Ferguson and Furth 1958, Brower *et al.* 1958) and in toroidal tubes (Conner *et al.* 1958). In most cases with high rate of rise of current a skin current is formed, as in the absence of B_z, and there is sufficient inertia in the contraction for some bouncing of the discharge to occur after maximum contraction (Kerst 1958, Sundstrom and Svennerstedt 1959). The reduced contraction means less compressional heating of the discharge and ohmic heating becomes an important contribution. Since the currents are confined to a sheath the ohmic heating causes the plasma pressure to build up mainly in this annular region.

In experiments with high magnetic fields in straight tubes (Kerst 1958), in spite of the high rate of rise of current, the discharge current was observed to be

approximately uniformly distributed across the tube even as early as 0·3 μsec after initiation. This may have been due to distortion of the axial magnetic field near the electrodes since copper electrodes were used and the axial magnetic field was applied as a short pulse.

3.2.3. *Low rate of rise of current, no axial magnetic field.*

When discharges are produced with low electric fields so that the rate of increase of current is slow compared with the skin depth penetration time, the initiation of the discharge is markedly different to that described above. Instead of the discharge forming in a sheath near the tube wall, the high-speed photographs show the discharge growing from the start as a constricted channel along the tube axis (Hemmings, Miles and Ware 1957). This channel grows in brightness with time and often expansion rather than contraction is observed (Allibone *et al.* 1958).

A possible explanation is as follows. Since the discharge remains a poor conductor for a considerable time during the slow ionization of the gas, the electric field will penetrate to the centre of the tube and, initially, ionization will occur fairly uniformly over the tube cross section. When the self-magnetic field becomes large enough to affect the electron motion it will inhibit further ionization in the outer regions of the discharge since this is where the magnetic field is largest. In the centre, where the magnetic field is small, the ionization can proceed unimpeded. Further, the growth of the central region will be enhanced by the pinch effect drawing the charged particles from the outer regions into the central core. The expansion of the channel may be due to ohmic heating increasing the plasma pressure faster than the growth of the magnetic pressure.

3.2.4. *Low rate of rise of current with an axial magnetic field.*

At fairly low currents (10^3–10^4 A in tubes 10–30 cm diameter) in the presence of an axial magnetic field the discharge forms over progressively more of the tube as the applied magnetic field is increased (Hemmings, Miles and Ware 1957, Allibone *et al.* 1958, Dolgov-Saveliev *et al.* 1958). With high fields the discharge fills most of the tube throughout the pulse. A notable feature of these discharges is the appearance of multiple channels on the streak photographs at the very beginning of the luminosity and continuing throughout the pulse in some cases (Hemmings, Miles and Ware 1957).

At higher currents ($\sim 10^5$ A) in the toroidal tubes ZETA and SCEPTRE III, the discharge forms over the whole tube, but as the current builds up the streak photographs show a fairly sharp-edged channel constricting to about half the tube diameter (Thonemann *et al.* 1958, Allen *et al.* 1958). This constriction occurs before B_θ becomes equal to the applied magnetic field B_{z0}. The intermittent light observed outside the main channel (Ware 1959b) and magnetic probe measurements (Allen and Liley 1959) indicate a considerable amount of plasma left behind by the constricting discharge. During or shortly after constriction the channel becomes invisible to normal high-speed photography, but its continued existence has been demonstrated using more sensitive detection in the ultra-violet spectrum (Allibone *et al.* 1958), as shown in Fig. 6. By measuring the current density and axial field at the centre of the discharge with magnetic probes, Lees and Rusbridge (1959) have inferred that the constriction ceases when the discharge

is about half the tube diameter, and that when the parameter $2I/bB_{z0}$ exceeds about 1·4, negative B_z is generated in the outer parts of the discharge.

High currents with low dI/dt in straight tubes, with very high pulsed magnetic fields (Golovin *et al.* 1958), showed discharges forming at the centre of the tube and then expanding as the current increased. The width of the channel at peak current was smaller the higher the magnetic field. Here again large copper electrodes were used and the magnetic field was pulsed, so that the constriction may have been caused by the non-uniformity of the magnetic field near the electrodes.

§ 4. Discharge Equilibrium

4.1. *Linear Discharges*

Strictly speaking, complete equilibrium for a discharge, with all the parameters remaining constant in time, can never be achieved. There must at least be some heat generation due to particle collisions. However, a steady state can be considered in which electric fields (or fusion reactions) provide the heat energy, and thermal conduction or radiation dissipate the heat. For discharge configurations which involve magnetic fields frozen into the plasma, the discharge is not even in a steady state. However, a quasi-steady state can be assumed for times short compared with the time taken by the magnetic field to diffuse out of the plasma.

The prime condition for equilibrium is that the gas has no acceleration, so that from Eqn (6)

$$\frac{1}{c}\mathbf{j}\wedge\mathbf{B} = \nabla nk(T_e + T_p). \qquad \ldots\ldots(22)$$

Even for a straight discharge with assumed cylindrical symmetry this equation together with Maxwell's equation (13) for $\mathbf{j}$ has four too many unknowns. Ideally, the heat flux and energy transfer equations must be solved simultaneously. Such detailed calculations have not been carried out. Not only are the equations complex but the temperature predicted by such solutions would greatly exceed the observed temperatures and would not be very relevant to present experimental results. The practice has been to make simplifying assumptions about the particle temperatures.

In the case where the self-field B_θ is the only component of $\mathbf{B}$ present, Eqn (22) has a simple integral giving the mean temperature (or mean pressure) in the discharge (see § 1, Eqn (2)), but this solution gives no information about the discharge configuration. In the original treatment by Bennett (1934) a plasma column was considered in which the electrons were assumed to have an axial drift velocity constant with respect to radius. With constant temperatures throughout the discharge this leads to the simple density distribution

$$n = n_0/(1 + Ar^2)^2 \qquad \ldots\ldots(23)$$

where A is a constant. Essentially, similar distributions were obtained by Tonks (1939), Thonemann and Cowhig (1951) and Blackman (1951). The effect of bremsstrahlung radiation on the equilibrium conditions for a highly constricted discharge was considered by Braginsky (1957 b) and Pease (1957), who showed

that the current in such a discharge in the absence of B_z could not exceed about 2×10^6 A.

When an axial magnetic field is also present and is assumed frozen into the plasma, there is an infinite number of quasi-steady state solutions depending on the relative magnitudes of B_θ and B_z and their gradients. The simplest configuration occurs when the discharge is initiated rapidly in the presence of a uniform B_z. In this case B_z is approximately constant within the discharge channel and falls rapidly at the edge to a low value outside. B_θ is zero inside and rises rapidly to a large value outside where it falls off with radius as $1/r$. The width of the transition layer depends on the conductivity of the plasma and the speed with which the discharge is generated. When this sheath thickness is small compared with the radius, Eqn (22) can be integrated over the sheath in a similar manner to Eqn (16), to give

$$\frac{B_{\theta e}^2}{8\pi} + \frac{B_{ze}^2}{8\pi} = \frac{B_{zi}^2}{8\pi} + p \qquad \ldots\ldots(24)$$

where the subscripts e and i refer to fields on the external and internal surfaces of the sheath, and where both $B_{\theta i}$ and the external plasma pressure are assumed zero.

In the case with applied B_z where a true steady state has been reached (i.e. $\partial \mathbf{B}/\partial t$ truly zero) current will still flow in the θ direction due to the anisotropic conductivity of an ionized gas in the presence of a magnetic field. This was first pointed out by Braginsky and Shafranov (see Bezbatchenko *et al.* 1956). The equilibrium solution for the case where the radial electron temperature gradient is assumed zero has been obtained by Bickerton (1958), and a similar result has been given by Longmire (1958).

4.2. *Toroidal Discharges*

The change to toroidal geometry introduces two main effects on discharge equilibrium. Firstly, since the B_θ magnetic field is larger on the concave side of the curved discharge than on the convex side, this component of field exerts a net outward force on the discharge tending to increase the torus diameter. A further component of magnetic field is required to give equilibrium. When trapped fields are not considered this must be a field vertical to the plane of the torus and of the required sign to exert an inward force on the toroidal current (Shafranov 1957, 1959, Biermann *et al.* 1957). Experimentally, if the discharge tube possesses a metal sheath the movement of the discharge towards the outer wall induces eddy currents which generate such a transverse field. The equilibrium position for the centre of the discharge is displaced outwards from the centre of the tubing an amount depending on the ratio of the torus and tube diameters. When a toroidal magnetic field parallel to the discharge is trapped in the plasma, this will reduce the tendency for the discharge to expand, but in general some externally applied field (e.g. due to eddy currents) is needed for equilibrium.

The other effect is connected with the toroidal magnetic field and the fact that its magnitude must be inversely proportional to R, where R is the radius of curvature of the lines of force. When this field component is not zero, Eqn (22) requires that $\nabla . \mathbf{j}_\perp \neq 0$, where $\mathbf{j}_\perp$ is the component of $\mathbf{j}$ perpendicular to $\mathbf{B}$ (Spitzer 1958). $\nabla . \mathbf{j}_\perp$ will have a different sign above and below the equatorial plane of

the torus. Since equilibrium requires that $\nabla . \mathbf{j}$ be zero, this will be possible only if currents can flow along the lines of force between regions of opposite sign for $\nabla . \mathbf{j}_\perp$. When there is a B_θ magnetic field present, as with the pinch discharge, the lines of force do pass from top to bottom of the torus and equilibrium is possible. When only j_θ currents, and hence only B_z, are present, the lines of magnetic force must be given a rotational transform by some other means. The conditions for this equilibrium, and in particular the limitation on the plasma pressure, have been considered by Spitzer (1958) and Kruskal and Kulsrud (1958).

Some of the equilibrium configurations which are possible in a torus when trapped fields are considered have been derived by Laing, Roberts and Whipple (1959), who have looked for configurations close to those observed in the ZETA experiment (Butt *et al.* 1958).

§ 5. Magnetohydrodynamic Stability

5.1. *Theory*

Many theoretical studies have been made on the stability of magnetically confined plasmas and in particular of the pinch discharge. In most cases the starting point has been the transport equations obtained by making the magnetohydrodynamic approximations referred to in § 2. Other treatments based on approximate solutions of the Boltzmann equation ignoring collisions (see § 2.3) have been made by Chandrasekhar, Kaufmann and Watson (1958), Rosenbluth and Longmire (1957), Kruskal and Oberman (1958) and Sagdeyev *et al.* (1958), but the results are in close agreement with magnetohydrodynamic solutions. One reason for this is the fact that to the degree of approximation to which the solutions have been taken the electric conduction equation is identical with the perfect conductivity approximation (Eqn (15)). Also it has been shown by Rosenbluth and Rostoker (1958) that to the same degree of approximation, if the magnitude of **B** does not vary along a line of force, as is the case in pinch discharge, the plasma momentum balance equation is the same as in the magnetohydrodynamic approximation. This is valid for marginally stable configurations, which are the ones studied in deciding stability criteria. The few treatments which have used a higher order approximation for the electric conduction equation are considered in § 5.3 below.

The conditions for stability are obtained by studying the behaviour of small perturbations to the discharge. Terms which are of second order or higher in the small increments to the discharge parameters are neglected so that a set of linear equations are obtained. Any arbitrary displacement can be regarded as a sum of Fourier components with time variation of the form $\exp i\omega t$, so that the increments in the various discharge parameters can be represented by expression of the type given in Eqn (4). Because of the linearity of the equations each Fourier component (or normal mode) can be treated independently. One procedure is therefore to substitute expression (4) into the plasma equations. Since all the terms have the same t, θ and z variation, the exponential terms cancel, and a series of linear simultaneous differential equations remain which must be solved for the radial dependence of the amplitudes $A'(r)$. Finally, by applying

the boundary condition, a dispersion relation results, giving ω as a function of k.

In general ω can be real, imaginary or complex, but when the magneto-hydrodynamic approximations are made, including the perfect conductivity approximation, a property of this particular set of equations is that ω^2 is always real. In the case when $\omega^2 > 0$, the solutions represent oscillations or waves with constant amplitude and the discharge is stable. (There are no dissipative effects in this set of equations so the oscillations are not damped.) When $\omega^2 < 0$ then ω can have a negative imaginary value and the perturbations will grow with time. The conditions for stability are therefore obtained by studying the dispersion relations in the neighbourhood of the marginal stability where $\omega = 0$.

An alternative treatment which avoids some of the complexity of the normal mode analysis and which yields stability conditions but no dispersion relation, is to evaluate the change in the electromagnetic and thermal energy of the system which results from an arbitrary displacement of the plasma. If the change in this 'potential energy' is positive for all displacements of the plasma, the configuration will be stable. This energy principle was first applied to a plasma by Lundquist (1951) and has been developed by Bernstein, Frieman *et al.* (1958) and Hain, Lust and Schluter (1957). The potential energy δW is obtained in the form of a volume integral which is a function of the discharge configuration and the displacement ξ undergone by the plasma, where ξ is initially an arbitrary function of position. The usual procedure is to find the displacement which minimizes δW. This will be the most unstable displacement and if δW is positive the discharge will be stable. For a given instability mode (m, k) provided

$$mB_\theta + krB_z \neq 0 \qquad \ldots\ldots(25)$$

or at most only vanishes at discrete radii, the energy principle shows that the most unstable displacements are those for which $\nabla . \xi = 0$, i.e. displacements involving no compression or expansion of the plasma. Most of the theoretical treatments of stability have studied such discharges. The exceptional case is considered in § 5.1.3 below.

In neither approach are large amplitude displacements considered, since they lead to intractable non-linear equations. Only on very simplified models of the discharge have such effects been considered (Roberts and Tayler 1957, Rosenbluth and Longmire 1957, Tayler 1959 c).

The following is a summary of the results of linear stability theory.

5.1.1. *Perfectly conducting compressible cylinder with infinitely thin current sheath.*

This simplified model, which gives the discharge a sharp edge and which assumes it to be a perfect electrical conductor, was introduced originally by Kruskal and Schwarzschild (1954) and was used in much of the early stability theory. With it Shafranov (1956), Tayler (1957 b) and Kruskal and Tuck (1958) showed that a magnetic field trapped in the discharge has a stabilizing effect but that modes with long wavelength in the z direction were still unstable. Complete stability is obtained when this discharge model is surrounded by a tube with conducting walls, as was shown by Shafranov (1956), Rosenbluth (1956) and Tayler (1957 b). Fig. 9 (taken from Tayler) shows the conditions for stability.

With increasing discharge compression the allowable values for the external B_z field and for the β value ($\beta = p/[B_{\theta e}^2/8\pi]$) decrease. For a compression greater than 25 ($b/a > 5$) there is no stability.

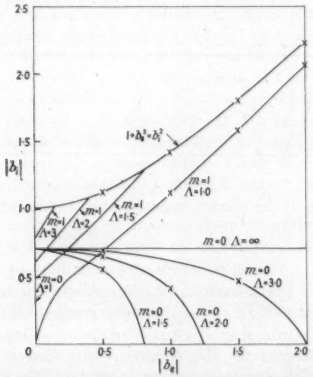


Fig. 9. Stability diagram for a perfectly conducting cylinder of radius a with surface currents and surrounded by a conducting wall at radius b. Stability occurs for a given Λ ($= b/a$) for points above both the $m = 0$ and $m = 1$ curves with this Λ and below the curve $1+\delta_e^2 = \delta_1^2$ where $\delta_e = B_{ze}/B_{\theta e}$ and $\delta_1 = B_{z1}/B_{\theta e}$. The quantity $\beta = 1+\delta_e^2-\delta_1^2$ (Tayler 1957 b).

5.1.2. *Perfectly conducting cylinder with finite current sheath.*

Using this more realistic model, Boon *et al.* (1958) estimated that for a compression factor of 4, instability would occur for current sheaths greater than 15% of the discharge radius. However, even within thin current sheaths there will always be a radius at which a particular instability mode (m, k) is parallel to the helical magnetic field lines so that

$$mB_\theta + krB_z = 0. \qquad \ldots\ldots(26)$$

Rosenbluth (1958) and Suydam (1958) showed that because of this singularity, instability will occur unless stringent conditions are satisfied by the magnetic field configuration. By solving the equations for marginal stability, Rosenbluth showed that only certain discharge configurations which have a negative B_z external to the discharge will be stable. The required variation of the field components through the current layer are not those which would arise from the internal and external fields diffusing into one another. However, provided the negative field is present and the degree of constriction limited, the instability growth rate in this last case will be very small (Drummond and Rosenbluth 1960). Suydam found a necessary but not sufficient condition for stability to be

$$\frac{r}{4}\left(\frac{1}{\mu}\frac{d\mu}{dr}\right)^2 \geqslant -\frac{8\pi}{B_z^2}\frac{dp}{dr} \qquad \ldots\ldots(27)$$

where $\mu = B_\theta/rB_z$, so that $2\pi/\mu$ is the pitch of the lines of force. The quantity in the brackets is a measure of the shear in the lines of force as one passes from one radius to another. For stability, therefore, the magnetic field must have sufficient shear.

These instabilities will be localized in a thin layer at the radius where (26) is satisfied, but it is unknown how they will affect the remainder of the discharge when their amplitude grows. Since they arise because of a singularity at a discrete radius, whereas the charged particles perform helical paths and experience the fields over a finite band of radii, Hubbard (1958) made an approximate calculation to estimate the effect of the finite Larmor radii of the particles, since this is neglected in the magnetohydrodynamic approximation. His results suggest that stability may still be achieved with discharges which have thin current layers and which satisfy the stability conditions for the infinitely thin layer model.

5.1.3. *Perfectly conducting discharge with constant pitch magnetic field.*

If $\partial(B_\theta/rB_z)/\partial r$ is zero over the whole discharge so that the helical lines of force have a constant pitch, there will be a set of m and k which will satisfy Eqn (26) at all radii. These instabilities are parallel to the magnetic field. In this case a helical line of force with the plasma frozen to it will move outwards (or inwards) at all points along its length and will accurately match the line of force it has displaced. In some cases little or no change in the magnitude of the magnetic field occurs, the energy for the instability coming from the expansion of the hot plasma. Such instabilities have been studied by Johnson *et al.* (1958), who refer to them as 'interchange instabilities', by Kadomtsev and Braginsky (1958), and by Kadomtsev (1959), who uses the equally appropriate name 'convective instabilities'.

The most unstable perturbations in the case of a convective instability satisfy the condition

$$\nabla . \boldsymbol{\xi} = 2B_\theta^2 \xi_r/r(B_\theta^2 + B_z^2 + \gamma p), \qquad \ldots\ldots(28)$$

as shown by Tayler (1959 b), where γ is the ratio of the specific heats and is equal to 5/3. The condition for stability is then

$$\frac{dp}{dr} > -2\gamma p B_\theta^2/r(B_\theta^2 + B_z^2 + \gamma p). \qquad \ldots\ldots(29)$$

Thus the pressure gradient must be less negative than the negative quantity on the right-hand side. Close to marginal stability these convective modes produce little change in the magnetic field (Ware 1961). However, it can be shown that these convective instabilities are not the most unstable modes for a constant pitch discharge. Other modes which are not parallel to **B** and which satisfy $\nabla . \boldsymbol{\xi} = 0$ require a positive pressure gradient for stability (Ware 1961).

5.1.4. *Discharges with $B_z \gg B_\theta$.*

In the case of infinitely thin skin currents with no external conducting tube, and where, both within the discharge and external to it, B_z is large compared with B_θ, the discharge is stable against all modes except $m = 1$. With $m = 1$ the discharge is unstable only for wavelengths greater than $2\pi a B_z/B_{\theta e}$ (Tayler 1957 b, Kruskal and Tuck 1958). a is the discharge radius.

This result is important in the case of a torus, since round the closed path of the torus there must be an integral number of instability wavelengths. The torus circumference L_T is an upper limit to the possible wavelength and for $m = 1$ the discharge will be stable if $L_T < 2\pi a B_z/B_\theta$ (Tayler 1957b, Braginsky and Shafranov 1958, Kruskal *et al.* 1958).

In the case of a torus which has a rotational transform angle ι due to some cause other than B_θ, the condition for stability becomes (Kruskal *et al.* 1958)

$$L \leqslant B_z(\iota + 2\pi h) r/B_\theta \qquad \ldots\ldots(30)$$

where h is the integer (positive or negative) which gives the right-hand side its smallest positive value. In this case, unless ι is an integral multiple of π, the criterion depends on the relative signs of B_θ and B_z. The values of B_θ are those which reduce the transform angle to 0 or increase it to 2π. When distributed currents are considered (Johnson *et al.* 1958) the same condition is obtained for stability against $m = 1$ modes. However, in this case higher m modes can be unstable and the corresponding stability conditions have been obtained.

5.1.5. *Discharge models with dissipative effects.*

Very little has been done on the affect of particle collisions on stability because of the increased complexity of the plasma equations. Viscosity can be expected to decrease the growth rates for short wavelength instabilities and this has been shown for one case by Tayler (1957a). The inclusion of finite conductivity is particularly difficult (Tayler 1959a).

5.1.6. *Stable configurations.*

Using the energy principle, Laing (1958) has constructed discharge configurations which were obviously stable by requiring that all the terms in the integral for δW be positive. The unusual features of these configurations are (*a*) a positive pressure gradient in the centre of the discharge, and (*b*) negative axial current in the outer regions so that the total axial current is zero. Such current configurations were observed by Burkhardt and Lovberg (1958a) at the current zeros in their oscillating discharge and the fluctuations at these times were small. Johnson *et al.* (1958) have shown that stable configurations can be produced in high B_z discharges with helical windings.

5.1.7. *Toroidal discharges.*

Several workers (Suydam *et al.* 1960, Rotenberg *et al.* 1960) are studying the difficult problem of analysing the instability of a toroidal discharge, but no conclusive results have yet been reached. In the case of small plasma pressure, Johnson (1960) has shown that the curvature does not affect the wriggle instability.

5.2. *Experimental Observations on Stability*

5.2.1. *Discharges with trapped axial magnetic field.*

For those discharges which are generated rapidly so that a skin effect occurs and the B_z field is trapped in the plasma, there is a qualitative agreement between

theory and experiment. Thus, if only a small B_z is used so that the discharge compresses to a small diameter, a large amplitude instability is observed (Tuck 1958, Colgate, Ferguson and Furth 1958) involving lateral movements of the discharge (Burkhardt and Lovberg 1958 b). With larger fields, however, these large amplitude fluctuations cease and only a small flutter is observed on the magnetic probe signals (Burkhardt and Lovberg 1958 a, Honsaker *et al.* 1958, Colgate, Ferguson and Furth 1958). If a thicker insulating tube is used between the discharge and the metal tube so that a larger B_z field is left outside the discharge, Colgate, Ferguson and Furth (1958) observed that the large amplitude instability occurred with less discharge compression. These results are in qualitative agreement with the conditions for gross stability referred to in § 5.1.1 above.

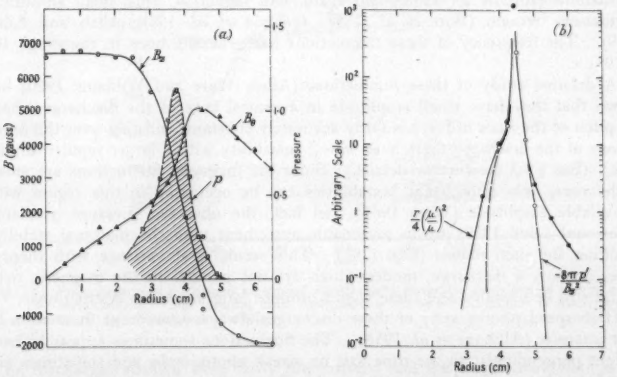


Fig. 10. (*a*) Radial distribution of measured B_θ and B_z, and the deduced plasma pressure, in the linear pinch discharge experiment COLUMBUS S-4 (Tuck 1959). (*b*) Plots of the experimental values of the two terms in the Suydam stability criterion against radius for the conditions shown in (*a*) (Tuck 1959).

It is an obvious inference that the remaining small amplitude fluctuations are the layer instabilities referred to in § 5.1.2, since in these experiments the current sheaths were thicker than those required for complete stability, even on the basis of Hubbard's calculations. However, a study of these fluctuations (Burkhardt and Lovberg 1958 b) has yielded little information about their structure and experimental proof that they are the layer instabilities is lacking. Typical mean frequencies for these fluctuations are in the range 10^6–10^7 c/s.

Attempts were made by Colgate, Ferguson and Furth (1958) and by Burkhardt and Lovberg (see Tuck 1959) to stabilize the discharge completely by programming the B_z field so that B_z became negative outside the discharge, but were unsuccessful. However, there is some evidence (Birdsall *et al.* 1959 b, Aitken, Burcham and Reynolds 1959) that the fluctuations observed in straight discharge tubes (as used in these experiments) are generated at the electrodes. An interesting result

obtained by Burkhardt and Lovberg (Tuck 1959) was that, as the magnetic shear in the discharge was increased by increasing the negative B_z, the pressure gradient also rose so as to satisfy approximately the equality in the Suydam criterion (Eqn (27)). Fig. 10(*a*) shows a radial plot of B_θ, B_z and p from the results, and Fig. 10(*b*) shows the two terms in the Suydam criterion. The pressure in the discharge is apparently increasing until the pressure gradient is sufficient for instability.

5.2.2. *Discharges with approximately constant pitch magnetic fields.*

In discharges with applied B_z and a low rate of current no skin effect is observed and the magnetic field is observed to have approximately the same pitch at all radii (Ware 1959c). With low B_z these discharges also exhibit large fluctuations (Butt *et al.* 1958), and again with higher B_z only small amplitude fluctuations remain (Butt *et al.* 1958, Golovin *et al.* 1958, Allen and Liley 1959). The frequency of these fluctuations has generally been in the range 10^4 to 10^5 c/s.

A detailed study of these fluctuations (Allen, Ware and Williams 1960) has shown that they have small amplitude in a central core of the discharge where the pitch of the lines of force is fairly accurately constant, and that over the outer regions of the discharge there is an $m = 1$ instability with a larger regular amplitude. (See § 5.3 for further details.) Since the magnetic fluctuations are small in the core, only interchange instabilities can be occurring in this region with appreciable amplitude (Ware 1961). In fact, the observed pressure gradients (Allen and Liley 1959) are in reasonable agreement with the marginal stability condition for such modes (Eqn (29)). This result is at variance with theory, since, in such a discharge, modes which are not parallel to the magnetic field should still be unstable and these would produce large magnetic fluctuations.

High-speed photography of these discharges always shows large fluctuation in light intensity (Allibone *et al.* 1958). The fluctuations sometimes appear as bars of light perpendicular to the time axis on streak photographs and sometimes are helical in shape. A close correlation has been observed between these light fluctuations and the magnetic fluctuations (Ware 1959b). (See also § 5.3.)

5.2.3. *Discharges with* $B_z \gg B_\theta$.

This type of discharge has been studied extensively at Princeton University and it is one of the few cases where close quantitative agreement has been found between stability theory and experiment (Kruskal *et al.* 1958). With increasing voltages applied to the discharge the current is found to increase rapidly with voltage until a critical current is reached, after which it increases very slowly with voltage (see Fig. 11). Above the critical current fluctuations are observed and large amounts of impurity appear. The magnitude of the critical current is different when the sign of the current is changed, and the values of these two critical currents are in good agreement with the two values of B_θ for the limits of stability given by Eqn (30).

At that time instability modes with higher values of m, which were expected because of the distributed discharge current, were not observed. These modes have since been detected by Bernstein, Kranz and Tenney (1959).

5.3. *Instability Waves*

As mentioned above, in most stability theory the approximation of perfect conductivity (Eqn (15)) has been used. The solutions for ω^2 are then always real. An instability configuration will remain fixed in position and merely grow in amplitude with no oscillations or wave motion occurring. Experimentally, however, oscillations are always observed in connection with instabilities. The motion associated with these oscillations could be a mass motion of the discharge or a wave motion. Of the many observations on instabilities, only two experiments, until recently, involved measurements capable of answering this question (T. K. Allen 1957, Baker, Sawyer and Stratton 1958). Both indicated wave motion.

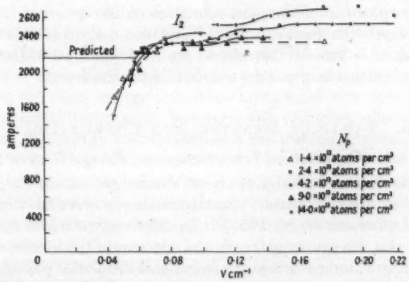


Fig. 11. Peak current as a function of applied electric field at various pressures of helium in the B-1 STELLARATOR. The Kruskal limit predicted for the $m = 1$ instability is shown by the horizontal line (Kruskal *et al.* 1958).

The discharges studied were fairly low current and the instabilities were $m = 1$ modes with observed wavelengths roughly equal to the tube diameter. The wave velocities in the z direction were in the range 10^5 to 10^6 cm sec^{-1}, being in the opposite direction to the gas current. Similar velocities have been observed with a growing $m = 1$ instability (see Allibone *et al.* 1958). Further properties of these waves have recently been studied by Lees *et al.* (1960).

It has been pointed out by Ware (1960) that if the Hall e.m.f. ($\mathbf{j} \wedge \mathbf{B}/ne$) is retained in the electric conduction equation (see Eqn (7)) such wave motion is expected. In this case the magnetic lines of force are frozen to the mean electron velocity (see § 2.5) and the instability perturbations to the magnetic field are carried along by the electrons. For the positive ions it is a wave motion. Ware estimated that the wave velocity should be approximately the mean electron velocity perpendicular to the unperturbed field, namely,

$$\mathbf{V}_w = \nabla p \wedge \mathbf{B}/neB^2. \qquad \ldots\ldots(31)$$

That such a wave velocity is obtained when the full set of plasma equations are solved with the Hall e.m.f. retained has been shown independently for the special case of an $m = 0$ mode in a discharge with $B_z = 0$ in the work of Kadomtsev

(1959) and Ridakov and Sagdeyev (1959). This wave velocity agrees in sign and order of magnitude with the observed wave velocity and has the correct dependence on magnetic field (Ware 1960).

A search for such waves in the high current toroidal discharge experiments, SCEPTRES III and IV, has revealed two such instability waves (Allen, Ware and Williams 1960). In the central core of the discharge, where dp/dr is negative, the fluctuations travel in the opposite direction to the discharge current; in the outer regions, where dp/dr is positive, the $m = 1$ instability observed in this region moves in the same direction as the current. These results agree in sign and order of magnitude with Eqn (31). (The possibility that two instabilities might occur simultaneously in different regions of the same discharge was suggested by Kadomtsev (1959).) Since in earlier work it has been shown that the 'bars' on the streak photographs exhibit similar velocities (Allibone *et al.* 1958) and that there is a close correlation between 'bars' and the magnetic fluctuations (Ware 1959b), the evidence is strong that the 'bars' are identical with the magneto-hydrodynamic instabilities in the outer regions of the discharge.

§6. Particle Heating and Energy Loss

6.1. *The Electron Temperature and Energy Loss*

A characteristic feature of high current discharges is the large amount of energy dissipated in the discharge; instantaneous powers as high as 900 MW have been quoted (Conner *et al.* 1958). Equally characteristic is the fact that although most of this energy is given to the electrons, the electron temperature remains low. Electron temperature measurements from the plasma conductivity (Artsimovich 1958, Colgate, Ferguson and Furth 1958, Tuck 1958, Allen and Liley 1959) and from spectroscopic observations (Kaufman and Williams 1958, 1960) have yielded values less than 5×10^5 deg K, whereas the energy input was sufficient to produce temperatures between one and two orders of magnitude greater. The electrons must be losing energy at roughly the same high rate as they receive it.

The electrons can lose energy at these rates either by exciting radiation from the bound electrons of impurity ions, if a sufficient concentration of impurity ions is present, or by the electrons transporting the energy to the walls. Calculations of the rates of radiation expected for given impurity concentration and electron temperatures have been made by Knorr (1958) assuming that the discharge pulse time is long compared with the ionization times for the multi-ionized ions concerned. These results have been applied by Kogan (1959). In connection with particle transport of energy, since the above temperature and energy measurements are for toroidal discharges, any such transport must be across the magnetic field. The energy flux due to normal diffusion and heat conduction across the magnetic field as given by Eqns (7) and (8) is much smaller than the observed energy loss, and hence if the electrons are transporting energy at this rate some extra mechanism must be involved to enable the particles to move across the magnetic field.

The first experiments to determine which of these two processes was the main energy loss were made by Colgate, Ferguson and Furth (1958) with their ceramic

toroidal discharge tube 'gamma pinch'. They found that for the first eighth of a cycle of the current (about 5 μsec) the main energy loss to the walls was by electrons with energies of about 1·5 kev. These electrons were detected by the soft x-rays they induced. After about 10 μsec the ultra-violet light increased greatly in amplitude and became the main loss from then on. During the second eighth of a cycle neither the loss of energetic electrons nor the radiation was sufficient to explain the energy loss. However, since the lower limit for soft x-ray detection was 1 kev, lower energy electrons could have been escaping undetected during this period.

In the toroidal experiment, PERHAPSATRON S4, which had a quartz torus, ultra-violet radiation became intense after about 7 μsec, and when integrated for the duration of the discharge was the main energy loss (Tuck 1959). With PERHAPSATRON S5, by improving the vacuum conditions, the onset of this radiation has been delayed to 40 μsec (Karr 1960). Few x-rays were detected during this period but low-energy electrons were not studied. In experiments in a straight porcelain tube Kirillov (1959) found the ultra-violet radiation due to impurities to be the main energy loss when integrated over the current pulse.

In the large metal torus ZETA, however, the radiation was found to account for only 10% of the energy loss throughout the discharge pulse (Gibson, Mason and Barsanti 1960). Soft x-ray emission from the tube walls was intense but only at the end of the current pulse when a rapid fall of current occurred. A large density of slow electrons was detected near the tube walls and was consistent with these electrons carrying most of the energy to the walls. In the smaller metal torus SCEPTRE III the energy loss by electrons with energy greater than 2 kev was found to be small, but large currents to the walls of lower energy electrons were detected (Ware 1959 b).

These various results can be reconciled if it is assumed that the fundamental loss process is an electron particle loss. In the case of ceramic tubes this leads to wall vaporization and contamination of the discharge, so that radiation then becomes the main energy loss mechanism. In the metal tubes the particle loss will cause a much smaller rise in wall temperature and hence less contamination.

It has been suggested by Ware (1959 c) that the instability waves (see § 5.3) provide a mechanism by which electrons can escape across the magnetic field. In a magnetohydrodynamic instability there are regions where both electrons and ions drift radially outwards. Since these regions are carried along with approximately the same velocity as the equilibrium mean electron velocity, the unfortunate electrons in such a region will remain in it and will continue to move outwards until they are lost to the walls. They will be replaced by cold electrons (from secondary emissions, arc spots or ionization of neutral gas emitted from the walls), which move into the discharge in those regions of the instability where the net movement is inwards. The positive ions, however, undergo an oscillating radial drift since the instability configuration passes over them, and their losses will be much less. Such a preferential electron loss is indicated by experiment since there is considerable evidence that the positive ion temperature exceeds the electron temperature.

Other instabilities which have been suggested as explanations of the electron particle loss are ion wave instabilities (Bernstein *et al.* 1960) and electron plasma

oscillations excited by runaway electrons (see, for example, Tuck 1958, Buneman 1958, Ellis, Goldberg and Gorman 1959). The ion wave instabilities will occur when the electron drift velocity becomes comparable with the thermal velocities in the plasma. (The exact critical drift velocity depends on the assumptions made about the electron and ion temperatures.) These waves will produce space charge fields and enhanced diffusion in the same manner as the instability waves referred to above, but there is as yet no evidence that they occur experimentally.

Runaway electrons will excite electron plasma oscillations if $dn'/dv > 0$ where $n'dv$ is the number of runaway electrons in the velocity range v to $v+dv$. As indicated above, in most of the toroidal pinch experiments high-energy electrons are not very numerous throughout the main part of the current pulse, an exception being the first eighth of a cycle in the gamma pinch torus. In ZETA only at the end of the pulse does the plasma density fall low enough for an appreciable number of runaway electrons to occur (Gibson, Mason and Barsanti 1960). In the stellerator experiments runaway electrons may be more important (Bernstein, Chen *et al.* 1958) and there is indirect evidence that they play a part in the loss of plasma from the discharge (Ellis, Goldberg and Gorman 1959).

6.2. *The Positive Ion Temperature and Heating Mechanism*

In the toroidal discharges ZETA and SCEPTRE III the spectral lines of highly ionized impurities show marked Doppler broadening (Thonemann *et al.* 1958, Allen *et al.* 1958). The shape of the line profiles, which are integrals for the long exposure times of many complete discharge pulses, corresponds to a Maxwellian velocity distribution (Allibone *et al.* 1958) and is independent of the direction in which the discharge is viewed (Hughes and Kaufman 1959). These results suggest that the broadening is due to thermal motion of the ions, the temperatures involved being in the range 10^6 to 7×10^6 °K. However, a turbulent mass motion of the plasma could give a Maxwellian velocity distribution when averaged over the long exposure times. Evidence against this has been obtained by Kaufman, Hughes and Williams (1960) who have compared the broadening of ions of different mass. Thus F VI and B IV, which have atomic masses 19 and 10·8 and which emit light from the same central region of the discharge, show different velocities but the same temperature.

If thermal motion is the cause of the broadening, since the energy exchange time between these impurity ions and deuterons is short compared with the current pulse, the deuteron temperature should be of the same order of magnitude. This is provided the rate of heating or cooling of either the deuterons or the impurities is not excessively large. The apparent high ion temperature is one of the paradoxes of the slow toroidal discharge, since it was expected that the electron temperature would always exceed the ion temperature because most of the ohmic heat goes to the electrons.

A study of the ohmic heating of impure deuterium has been made by Ware and Wesson (1960). In the case where a cooling mechanism keeps the electron temperature low, the presence of highly ionized impurity ions leads to a larger fraction of the ohmic heat being given directly to the deuterons. With a 14% concentration of O V the deuterons receive 7% of the heat when their temperature

Vol. 24
1961

is 2×10^6 °K. This rate of heating is sufficient to explain the observed ion temperatures, but an attempt at experimental verification proved inconclusive.

Other heating mechanisms which have been suggested are the instability waves (Ware 1959c) and magnetohydrodynamic waves (Reagan 1959, 1960).

§7. Nuclear Reactions

The two nuclear reactions which can occur between deuterons and which occur with approximately equal probability are :

(i) $D + D \rightarrow (^3He + 0{\cdot}8 \text{ Mev}) + (n + 2{\cdot}45 \text{ Mev})$

(ii) $D + D \rightarrow (T + 1{\cdot}0 \text{ Mev}) + (p + 3{\cdot}0 \text{ Mev})$.

The cross sections for these reactions increase rapidly with energy up to about 100 kev (Post 1956) and since the population of energetic particles increases exponentially with temperature the reaction rate is a particularly rapidly varying function of temperature (Post 1956, Thompson 1957).

These reactions can be studied by detecting any one of the four reaction products, the neutrons being the easiest to detect since they penetrate the tube walls. The reaction products are emitted isotropically with respect to the centre of mass of the reacting particles with the energies indicated in the above equations. Since in a hot gas the centre of mass of pairs of colliding particles will have a Maxwellian velocity distribution, the observed energy spectrum of a particular type of reaction product will show a broadening about the appropriate energy indicated above, but no shift.

If, however, reactions are caused by a directed beam of deuterons colliding with comparatively stationary deuterons, the particles emitted parallel to the beam will have a higher energy in the laboratory frame of reference, and those emitted in the reverse direction a lower energy than the above value. The measured difference in energy for the two directions can be used to calculate the mean centre-of-mass velocity of the reacting particles. This measurement is therefore one of the simple but not conclusive tests applied to determine if the reactions are thermonuclear in origin. Until very high temperatures and copious reactions occur, very stringent tests are needed before the origin of observed reactions can be ascertained conclusively (see Tuck 1958).

7.1. *The Linear Pinch without an Axial Magnetic Field*

Only in straight pinch discharges have neutrons been detected when no stabilizing magnetic field was used. (The rates of rise of current have generally been lower with toroidal discharges.) The reactions occur in bursts of short duration usually at the second time, but sometimes at the first or subsequent times of maximum constriction. Up to 3×10^9 neutrons per discharge pulse have been observed (Andrianov *et al.* 1958). The yield increases with current but at the highest currents shows a saturation and sometimes a decrease (Artsimovich, Andrianov, Dobrokhotov *et al.* 1956, Anderson *et al.* 1957). The yield is quenched by small percentages of impurity and greatly reduced by small applied axial magnetic fields. A study of the energy of the neutrons has invariably shown a large asymmetry indicating mean centre-of-mass velocities in the range 7×10^7 to

$1{\cdot}6 \times 10^8$ cm sec^{-1} parallel to the gas current. If the reactions are due to accelerated particles colliding with stationary ones, the corresponding energies of the former would be 20–100 kev (Anderson *et al.* 1957, Andrianov *et al.* 1958).

The most plausible explanation of these reactions is the theory of Anderson *et al.* (1957) which is based on the large inductive voltages associated with the necking off of the discharge at various points along its length due to the $m = 0$ instability. Among the evidence in favour of this theory are the large voltage pulses which are sometimes detected simultaneously with the neutron pulse (Anderson *et al.* 1957, Siegbahn and Ohlin 1958). The presence of high voltages in the discharge is indicated also by the generation of high-energy electrons with energies up to 300 kev (Podgornyi *et al.* 1958), and recently the accelerated deuterons have been detected (Hartman, Colgate and Munger 1960).

In the experiments of Curran *et al.* (1958), when the external circuit inductance was made very small so that dI/dt was very large, smaller neutron yields were detected. The yields, however, were much less sensitive to both impurity and an applied axial magnetic field than in the experiments described above. These results suggest that a different mechanism is inducing the nuclear reactions in this case, although similar energy shifts have been detected in the neutron energies (Bodin, Fitch and Peacock 1960). Other workers (Hagerman and Mather 1958, Siegbahn and Ohlin 1958) subsequently found that after an initial rapid drop in neutron yield with increasing magnetic field, a small yield remained which was relatively insensitive to magnetic field and impurity. Colgate (1958 a) has suggested that the two reaction-inducing mechanisms are the $m = 0$ and $m = 1$ instabilities, the $m = 0$ case being the mechanism sensitive to magnetic field and impurities. He suggested that in the experiments of Curran *et al.* (1958) the low circuit inductance leads to a broad pinch boundary (because the electric field near the walls stays high for a longer time) and that this favours the $m = 1$ instability.

7.2. *The Linear Pinch with an Axial Magnetic Field*

As mentioned above, nuclear reactions continue to occur in the presence of an axial magnetic field but at a lower rate. The reaction rate varies with discharge conditions in a manner markedly different to the case where $B_z = 0$. Thus the neutron yield shows only a slow decrease with increase in B_z and is relatively insensitive to impurity. In addition, the neutron burst lasts for a much longer time (Hagerman and Mather 1958) and is often centred at the current maximum (Burkhardt and Lovberg 1958 a). It is not correlated with the 'bouncing' of the discharge.

Mather and Williams (1958) have found that the neutron energy shift is much smaller than in the $B_z = 0$ case, and their results show that the reactions cannot be produced by a simple axial acceleration mechanism since the necessary deuteron current would exceed the observed discharge current. No satisfactory mechanism has yet been proposed to explain the nuclear reactions.

In experiments in which a pulsed B_z field was applied simultaneously with the gas current, Colgate *et al.* (1958) detected a short neutron burst occurring simultaneously with the onset of an $m = 1$ instability. No detailed mechanism has yet been proposed whereby the $m = 1$ instability can cause particle acceleration.

7.3. *The Toroidal Pinch with Magnetic Field*

The properties of the neutron yields obtained with toroidal discharges are similar to those for straight tubes with applied magnetic field. The one major exception is the variation with magnetic field. No reactions have been observed with zero magnetic field. A small field is needed before a detectable yield is obtained and beyond this point the yield increases with magnetic field (Colgate, Ferguson and Furth 1958, Allibone *et al.* 1958). In most cases the yield passes through a maximum and decreases at high fields (Harding *et al.* 1958, Conner *et al.* 1958, Ware 1959b). The emission period extends throughout most of the current pulse—for as long as 2 msec in the case of ZETA (Thonemann *et al.* 1958)—and sometimes into the second half-cycle (Allibone *et al.* 1958). The reactions occur in the central core of the discharge (Harding *et al.* 1958, Conner *et al.* 1958). Although there is evidence for high temperatures in these discharges (see § 6.2) the observed reaction rates exceed the expected thermonuclear reaction rates in many cases (Thonemann *et al.* 1958, Allibone *et al.* 1958).

Energy shifts similar to those for the linear pinch with applied field have been detected for the neutrons (Rose, Taylor and Wood 1958, Conner *et al.* 1958) and for the protons (Jones *et al.* 1958). The mean velocity for the centre of mass of the reacting particles is about 5×10^7 cm sec^{-1}. Evidence has also been obtained indicating the mean centre-of-mass velocity has a θ component of about 2×10^7 cm sec^{-1} (Hunt 1960). Such a component is expected since high-energy deuterons confined in the discharge should move approximately along the helical lines of force. That the energy shift is not due to a simple mass motion of a thermonuclear-reacting gas has been shown by the Doppler shift measurements of Hughes and Kaufman (1959). However, if a small fraction of the deuterons has a high temperature and is moving through the remaining deuterons with a velocity of the order 10^7 cm sec^{-1}, this would give the observed energy shift (Herdan and Hughes 1961). Hunt (1960) has analysed the observed proton spectra making the assumption that all thermal motion of the deuterons can be neglected and that the reactions are caused by accelerated deuterons colliding with stationary deuterons. The results indicate a wide spread of deuteron energies, the population increasing with decreasing deuteron energy.

§ 8. High Current Discharges with other Geometrical Configurations

High current discharges can be produced with an infinite variety of shapes. Not only can the shape of the discharge tube be varied, but also the geometry of the electric and magnetic fields. Lastly, the ingenious experimenter has the fourth dimension at his disposal since both of the fields can be 'programmed', i.e. varied in time. In many cases both theory and experiment are tending towards more complex systems in order to achieve a stable discharge. Here only the more widely studied discharge configurations will be reviewed.

8.1. *Azimuthal Discharges*

As mentioned in the Introduction, discharges with currents induced in cylindrical discharge tubes in the θ direction are now being widely studied. The

standard practice is to use a low-inductance coil wound on the tube and fed from a low-inductance condenser bank. The high current in the coil is allowed to ring. The rapidly rising axial magnetic field induces a θ current in the gas. Generally, only a poor degree of ionization is achieved in the first half-cycle although a contracting shock wave and 'bounces' are sometimes observed (Bodin *et al.* 1959 b). This is due to the axial magnetic field limiting the energy which one electron can gain between collisions to $2mc^2E^2/B^2$. By the end of the first half-cycle there is a sufficient plasma density so that space charge forces can counteract the limiting effect of the magnetic field. In the second half-cycle, therefore, large gas currents are induced at the periphery of the discharge, and as in the case of the pinch discharge with high dI/dt, a contracting shock wave is generated. This is shown by A in Fig. 12, which is taken from Kolb (1958). B is the boundary between the plasma and the magnetic field.

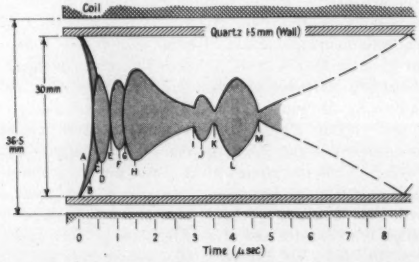


Fig. 12. Diagrammatic representation of the streak photograph for the second half-cycle of an azimuthal discharge with magnetic mirror ends (Kolb 1958).

As in the pinch discharge a series of bounces then occur (C, D, E, F, G in Fig. 12). If a uniform coil is used a damped train of many oscillations is observed (Bodin *et al.* 1959 b) but the plasma is rapidly lost longitudinally along the magnetic lines of force. To avoid this, most workers have designed their coils to give the magnetic field a larger magnitude at each end so that magnetic mirrors are formed. (For the principle of a magnetic mirror see Post 1958.) These mirrors greatly reduce the end losses, and in addition the radial components of magnetic field at these points exert a longitudinal force on the θ currents and cause longitudinal shock waves which travel towards the centre of the coil. These waves cause an expansion of the plasma cylinder to about half the tube diameter (H in Fig. 12). Subsequently the rising magnetic field compresses the discharge again to a small diameter. Then, as indicated at I, J, K, L, M, Kolb has observed expansions and contractions of the plasma cylinder suggesting some form of discharge oscillation. In some cases these oscillations have the appearance of a rotating instability (Kolb, Griem and Faust 1959, Bodin *et al.* 1959 a). In Kolb's experiment the visible part of these oscillations on the streak photographs does not expand to the walls. Other workers, however, have observed instabilities which do appear to

extend to the walls (Alidieres *et al.* 1959, Bodin *et al.* 1959 a, Kvartzhava, Kervalidze and Gvaladze 1959). In all cases the instability occurs in the second half-cycle.

Another important feature associated with the second half-cycle is the magnetic field within the plasma. The field produced by the coil in the second half-cycle will be defined as a positive field. Magnetic probe measurements (Kolb, Dobbie and Griem 1959, Fay, Hintz and Jordan 1959) have shown that some of the negative flux which penetrated the plasma in the first half-cycle remains trapped in the centre of the discharge during part of the second half-cycle, and then the field abruptly changes to a positive value.

Still further features associated with this interesting second half-cycle are the emission of neutrons and of soft x-rays from the plasma. In experiments with small or unity magnetic mirror ratios the neutron emission occurs in a short pulse (Kolb, Griem and Faust 1959, Fay, Hintz and Jordan 1959, Bodin *et al.* 1959 a) and in some cases there is correlation with the sudden change of trapped magnetic field. When strong magnetic mirrors are used, however, much larger neutron pulses of longer duration are observed (Elmore, Little and Quinn 1958, Tuck 1959, Kolb, Griem and Faust 1959). The SCYLLA experiment has yielded most of the latter results and in this experiment very extensive studies have been made of the reaction products. The neutrons show no energy shift (Elmore, Little and Quinn 1958), and accurate measurements of the protons and tritons have revealed a Maxwellian distribution for the centre-of-mass velocities of the reacting pairs of deuterons (Ribe 1959, Nagle *et al.* 1959). This is strong evidence that the deuteron motion is close to thermal motion in this experiment. The indicated temperature is about $1{\cdot}3 \times 10^7$ °K. The soft x-ray emission (Griem, Kolb and Faust 1959, Boyer *et al.* 1959) has a spectrum consistent with bremsstrahlung radiation† corresponding to an electron temperature of about 2×10^6 °K (Ribe 1959).

The heating mechanism causing the high ion temperature has yet to be determined. The shock wave and subsequent adiabatic compression will make substantial contributions, but the experimental results suggest that the instability is playing a major part. Thus the neutron production can be advanced to the first half-cycle provided the gas is adequately pre-ionized and provided an initial magnetic field is applied to the gas in the opposite direction to that generated by the coil in the first half-cycle (Kolb, Dobbie and Griem 1959, Quinn, Little and Ribe 1960). It has been pointed out by Furth (see Kolb, Dobbie and Griem 1959) that the presence of the trapped reverse field leads to an unstable configuration and has suggested the rapid dissipation of the reverse field as the cause of the heating. An attractive feature is that the instability changes the discharge to a more stable configuration and the particles heated by the instability will be contained.

8.2. *Annular Discharges* ('*Triax*' *and* '*Hard-core*')

At the Lawrence Radiation Laboratory some interesting annular discharges have been developed. These discharges are generated in the annular space between two concentric insulating cylinders of different diameter. The discharge is thus a hollow cylinder. By means of a copper conductor within the inner insulating tube a negative B_θ magnetic field is produced within the discharge which keeps the plasma from touching the inner tube.

† See footnote on page 32.

In the 'triax' experiment (Anderson, Baker, Ise *et al.* 1958) the current in the gas is larger than in the inner copper conductor and, hence, a positive B_θ is generated on the outer surface of the discharge and this keeps the plasma from the outer wall. No axial field is used. The two B_θ fields compress the plasma to a thin sheet and some 'bouncing' is observed. Such a system should be more stable than a normal pinch discharge without an axial magnetic field but not completely stable. Experimentally the discharge is stable for a longer time but an instability sets in which is accompanied by an increase in gas resistance and nuclear reactions. It was possible to measure the neutron energies with only limited accuracy but the results showed that any energy shift must be less than that which would be caused by 3 kev accelerated deuterons reacting with stationary deuterons (Pyle, private communication). As pointed out by Furth (see Kolb, Dobbie and Griem 1959) the magnetohydrodynamic configuration in this discharge, in which two parallel magnetic fields of opposite sign are separated by a thin sheet of plasma, has many similarities with the second half-cycle of the azimuthal discharge. The mechanisms for deuteron heating or acceleration may be the same.

In the 'hard-core' experiment (Birdsall, Colgate and Furth 1959a) the current in the gas is approximately equal to the current in the inner conductor, so that there is little or no B_θ outside the discharge. The plasma is compressed from the outside by a B_z magnetic field. This configuration is very attractive since the plasma is on the outside of the curved lines of force. Many authors have shown that such a configuration should be stable (see, for example, Berkowitz, Grad and Rubin 1958). Experimentally, some of the discharges with this general configuration are free from magnetic field fluctuations whereas others are not (Birdsall, Colgate and Furth 1959a). The unstable discharges are those with large axial current densities. However, there is evidence that these instabilities form initially at the cathode and propagate along the tube (Birdsall, Colgate and Furth 1959b, Aitken, Burcham and Reynolds 1959). Similar discharges with the B_z field missing, and which are therefore not in equilibrium, were studied by Anderson, Furth *et al.* (1958).

Because of the electrode effects on stability and because even with a stable discharge large heat losses will occur to the electrodes, a toroidal version of this type of discharge has been proposed by Colgate and Furth (1958). The results obtained with this experiment, which is at present under construction (Birdsall, Colgate and Furth 1959b), should prove very interesting.

8.3. *Rotating Plasmas*

Since a plasma is an electrical conductor many of the electromagnetic processes which can be applied to a piece of copper can also be applied to a plasma. In particular, a discharge can be made to rotate in a manner analogous with the Faraday disk or with the rotor of a homopolar motor. The rotation is generated in an annular discharge tube by creating a radial current density j_r between coaxial electrodes in the presence of an axial magnetic field B_z. The force $j_r B_z$ causes the angular acceleration and magnetic pressure balances the centrifugal force. Such rotating plasmas with circumferential velocities up to 6×10^6 cm sec^{-1} were generated in the HOMOPOLAR I experiment (Anderson, Baker, Bratenahl *et al.*

1958). Because of the energy of the rotating plasma, the discharge acts like a condenser containing a medium of very high dielectric constant of 10^6 to 10^7 (Anderson *et al.* 1959).

The centrifugal force can be used to aid the confinement of a plasma in experiments aimed at thermonuclear reactions. In the experiment 'ixion' the rotating plasma is produced in a magnetic mirror field (Boyer *et al.* 1958) and in HOMOPOLAR III curved lines of force are used so that the centrifugal force will prevent the loss of particles along the magnetic field (Colgate 1958 b). Instabilities have been observed (Baker *et al.* 1959), and short circuiting of the radial electric field by breakdown at the surface of the end insulators is suspected (Baker *et al.* 1959, Baker and Hammel 1960).

8.4. *Plasma 'Guns'*

By suitably designing the leads to a high current discharge the magnetic field on one side of the discharge can be made to exceed greatly the field on the other side. This will cause an acceleration of the plasma in the transverse direction. If there are no material obstacles, the plasma will often break loose from the electrodes and be ejected with a high velocity (Bostick 1956, Kolb 1957, Scott *et al.* 1958, Finkelstein, Sawyer and Stratton 1958). The ejected volume of plasma often contains circulating currents and trapped magnetic fields, and has been termed a 'plasmoid' by Bostick (1956). Growing interest has recently been shown in such plasma guns because of their possible application to propulsion in outer space (von Engel 1959) and to the injection of energetic plasmas into confining magnetic fields in thermonuclear research (Tuck 1959).

More efficient acceleration can be achieved by using a pair of parallel rails as the electrodes. In this way the moving plasma stays in contact with the electrodes for a longer time and since a current continues to flow in the plasma, there is a longer period of acceleration and higher velocities are achieved for a given energy input. Such 'rail guns' have been studied experimentally by Artsimovich *et al.* (1957), Bostick (1958) and theoretically by Morozov (1957). Similar propulsion was achieved by Marshall (1958) using a moving ring discharge which was achieved by a travelling magnetic wave. More recently, rail guns with coaxial electrodes have been studied (Dattner 1959, Tuck 1959). Overall efficiencies from 30 to 40% have been achieved in the conversion of stored energy to plasma energy (Tuck 1959, Marshall 1960, Bostick, Byfield and Nankivell 1960), involving several litres of plasma with density 10^{16} cm^{-3} and velocities up to $1{\cdot}5 \times 10^7$ cm sec^{-1}. The highest plasma velocities (4×10^7 cm sec^{-1}) have been obtained in the axial flow from a conical pinch discharge (Hartman, Colgate and Munger 1960).

Acknowledgments

The author wishes to thank the scientists whose photographs and diagrams are shown in this article for their assistance in reproducing these figures, and gratefully acknowledges the permission to use these illustrations granted by the scientists and by the North-Holland Publishing Company, the United Nations, the Institution of Electrical Engineers and the Editors of *The Journal of Applied Physics* and *Engineering*.

References

Since a large number of the following references are taken from the proceedings of the international conferences at Venice (1957), Geneva (1958), and Uppsala (1959), the following abbreviations will be made. *Terzo Congresso Internazionale sui Fenomeni D'Ionizzazione nei Gas, Venice* will be abbreviated to *Venice*; *The Proceedings of the Second International Conference on the Peaceful Uses of Atomic Energy, Geneva*, will be abbreviated to *Geneva*; and *The Proceedings of the Fourth International Conference on Ionization Phenomena in Gases, Uppsala*, will be abbreviated to *Uppsala*.

Aitken, K. L., Burcham, J. N., and Reynolds, P., 1959, *Uppsala*, **2**, 896.

Alfvén, H., 1950, *Cosmical Electrodynamics* (Oxford : Clarendon Press), p. 65.

Alidieres, M., Aymar, R., Etievant, C., Jourdan, P., and Samain, A., 1959, *Uppsala*, **2**, 1042.

Allen, J. E., 1957, *Proc. Phys. Soc.* B, **70**, 24.

Allen, J. E., and Reynolds, P., 1957, *Venice*, 33.

Allen, K. W., Bodin, H. A. B., Curran, S. C., and Fitch, R. A., 1957, *Venice*, 26.

Allen, N. L., Allibone, T. E., Chick, D. R., Hemmings, R. F., Hughes, T. P., Kaufman, S., Liley, B. S., Mack, J. G., Miles, H. T., Payne, R. M., Read, J. E., Ware, A. A., Wesson, J. A., and Williams, R. V., 1958, *Nature, Lond.*, **181**, 222.

Allen, N. L., and Liley, B. S., 1959, *Uppsala*, **2**, 937.

Allen, N. L., Ware, A. A., and Williams, R. V., 1960, *Bull. Amer. Phys. Soc.*, **5**, 352.

Allen, T. K., 1957, *Venice*, 40.

Allibone, T. E., Chick, D. R., Thomson, G. P., and Ware, A. A., 1958, *Geneva*, **32**, 169.

Anderson, O. A., Baker, W. R., Bratenahl, A., Furth, H. P., Ise, J., Kunkel, W. B., and Stone, J. M., 1958, *Geneva*, **32**, 155.

Anderson, O. A., Baker, W. R., Bratenahl, A., Furth, H. P., and Kunkel, W. B., 1959, *J. Appl. Phys.*, **30**, 188.

Anderson, O. A., Baker, W. R., Colgate, S. A., Furth, H. P., Ise, J., Pyle, R. V., and Wright, R. E., 1958, *Phys. Rev.*, **109**, 612.

Anderson, O. A., Baker, W. R., Colgate, S. A., Ise, J., and Pyle, R. V., 1957, *Venice*, 62 (see also 1958, *Phys. Rev.*, **110**, 1375).

Anderson, O. A., Baker, W. R., Ise, J., Kunkel, W. B., Pyle, R. V., and Stone, J. M., 1958, *Geneva*, **32**, 150.

Anderson, O. A., Furth, H. P., Stone, J. M., and Wright, R. E., 1958, *Phys. Fluids*, **1**, 489.

Andrianov, A. M., Bazilevskaia, O. A., Braginsky, S. I., Brezhnev, B. G., Khvaschevski, S., Khrabrov, V. A., Kovalski, N. G., Filippov, N. V., Filippova, T. I., Palchikov, V. E., Podgornyi, I. M., Prokhorov, Yu. G., and Sulkovskaya, M. M., 1958, *Geneva*, **31**, 348.

Artsimovich, L. A., 1958, *Geneva*, **31**, 6.

Artsimovich, L. A., Andrianov, A. M., Bazilevskaya, O. A., Prokhorov, Yu. G., and Filippov, N. V., 1956, *Atomnaya Energiya*, **1**, 76.

Artsimovich, L. A., Andrianov, A. M., Dokrokhotov, E. I., Lukyanov, S. Yu., Podgornyi, I. M., Sinitsyn, V. I., and Filippov, N. V., 1956, *Atomnaya Energiya*, **1**, 84.

Artsimovich, L. A., Lukyanov, S. Yu., Podgornyi, I. M., and Chuvatin, S. A., 1957, *J. Exp. Theor. Phys.*, **33**, 3.

Baker, D. A., and Hammel, J. E., 1960, *Bull. Amer. Phys. Soc.*, **5**, 341.

Baker, D. A., Sawyer, G. A., and Stratton, T. F., 1958, *Geneva*, **32**, 34 (see also Sawyer, G. A., *et al.* 1959, *Phys. Fluids*, **2**, 47).

Baker, W. R., Bratenahl, A., De Silva, A. W., and Kunkel, W. B., 1959, *Uppsala*, **2**, 1171.

Bennett, W. H., 1934, *Phys. Rev.*, **45**, 890.

Berger, J. M., Bernstein, I. B., Frieman, E. A., and Kulsrud, R. M., 1958, *Geneva*, **32**, 197.

Berkowitz, J., Grad, H., and Rubin, H., 1958, *Geneva*, **31**, 177.

Bernstein, I. B., Frieman, E. A., Kruskal, M. D., and Kulsrud, R. M., 1958, *Proc. Roy. Soc.* A, **244**, 17.

Bernstein, I. B., Frieman, E. A., Kulsrud, R. M., and Rosenbluth, M. N., 1960, *Phys. Fluids*, **3**, 136.

BERNSTEIN, W., CHEN, F. F., HEALD, M. A., and KRANZ, A. Z., 1958, *Geneva*, **32**, 210.
BERNSTEIN, W., KRANZ, A. Z., and TENNEY, F., 1959, *Phys. Fluids*, **2**, 713.
BEZBATCHENKO, A. L., GOLOVIN, I. N., IVANOV, D. P., KILRILLOV, V. D., and YAVLINSKY, N. A., 1956, *Atomnaya Energiya*, **1**, 26.
BICKERTON, R. J., 1957, *Venice*, 101 ; 1958, *Proc. Phys. Soc.*, **72**, 618.
BICKERTON, R. J., and LONDON, H., 1958, *Proc. Phys. Soc.*, **72**, 116.
BIERMANN, L., HAIN, K., JURGENS, K., and LUST, R., 1957, *Z. Naturf.*, **12a**, 826.
BIRDSALL, D. H., COLGATE, S. A., and FURTH, H. P., 1959 a, *Uppsala*, **2**, 888 ; 1959 b, *Uppsala*, **2**, 892.
BISHOP, A. S., 1958, *Project Sherwood* (Reading, Mass. : Addison-Wesley).
BLACKMAN, M., 1951, *Proc. Phys. Soc.*, **64**, 1039.
BODIN, H. A. B., FITCH, R. A., and PEACOCK, N. J., 1960, *J. Nucl. Energy*, part C, **1**, 206.
BODIN, H. A. B., GREEN, T. S., NIBLETT, G. B. F., and PEACOCK, N. J., 1959 a, *Uppsala*, **2**, 1061 ; 1959 b, *Uppsala*, **2**, 1065.
BODIN, H. A. B., PEACOCK, N. J., and REYNOLDS, J. A., 1959, *Uppsala*, **2**, 861.
BOON, M. H., LAING, E. W., RANDLES, J., ROBERTS, S. J., and TAYLER, R. J., 1958, *Atomic Energy Research Establishment, Report* T/R 2503.
BORZUNOV, N. A., KOGAN, V. I., and ORLINSKY, D. V., 1958, *Atomnaya Energiya*, **4**, 180.
BORZUNOV, N. A., ORLINSKY, D. V., and OSOVETS, S. M., 1958, *Atomnaya Energiya*, **4**, 149.
BOSTICK, W. H., 1956, *Phys. Rev.*, **104**, 292 ; 1958, *Geneva*, **32**, 427.
BOSTICK, W. H., BYFIELD, H., and NANKIVELL, J., 1960, *Bull. Amer. Phys. Soc.*, **5**, 366.
BOSTICK, W. H., LEVINE, M. A., and COMBES, L. S., 1953, *Phys. Rev.*, **89**, 339 (see also 1955, *Phys. Rev.*, **97**, 13).
BOYER, K., HAMMEL, J. E., LONGMIRE, C. L., NAGLE, D., RIBE, F. L., and RIESENFELD, W. B., 1958, *Geneva*, **31**, 319.
BOYER, K., LITTLE, E. M., QUINN, Q. E., SAWYER, G. A., and STRATTON, T. F., 1959, *Phys. Rev. Letters*, **2**, 279.
BRAGINSKY, S. I., 1957 a, *J. Exp. Theor. Phys.*, **33**, 459 ; 1957 b, *J. Exp. Theor. Phys.*, **33**, 645.
BRAGINSKY, S. I., and MIGDAL, A. B., 1958, *Acad. Sci. U.S.S.R.*, **2**, 20 (see also Andrianov *et al.* 1958).
BRAGINSKY, S. I., and SHAFRANOV, V. D., 1958, *Geneva*, **31**, 43.
BROWER, D. F., DUNAWAY, R. E., MALMBERG, J. H., OXLEY, C. L., STEARNS, M., KERST, D. W., SCOTT, F. R., CUNNINGHAM, S. P., and TUCKFIELD, R. G., 1958, *Geneva*, **32**, 110.
BUNEMAN, O., 1958, *Phys. Rev. Letters*, **1**, 8.
BURKHARDT, L. C., DUNAWAY, R. E., MATHER, J. W., PHILLIPS, J. A., SAWYER, G. A., STRATTON, T. F., STOVALL, E. J., and TUCK, J. L., 1957, *J. Appl. Phys.*, **28**, 519.
BURKHARDT, L. C., and LOVBERG, R. H., 1958 a, *Nature, Lond.*, **181**, 228 ; 1958 b, *Geneva*, **32**, 29.
BURKHARDT, L. C., LOVBERG, R. H., and PHILLIPS, J. A., 1958, *Nature, Lond.*, **181**, 224.
BUTT, E. P., CARRUTHERS, R., MITCHELL, J. T. D., PEASE, P. C., BIRD, M. A., BLEARS, J., and HARTILL, E. R., 1958, *Geneva*, **32**, 42.
CARRUTHERS, R., and DAVENPORT, P. A., 1957, *Proc. Phys. Soc.* B, **70**, 49.
CHANDRASEKHAR, S., KAUFMAN, A., and WATSON, K. M., 1957, *Ann. Phys., N.Y.*, **2**, 435 ; 1958, *Proc. Roy. Soc.* A, **245**, 435.
CHAPMAN, S., 1954, *Astrophys. J.*, **120**, 151.
CHAPMAN, S., and COWLING, T. G., 1939, *The Mathematical Theory of Non-uniform Gases* (Cambridge: University Press).
CHEW, G. F., GOLDBERGER, M. L., and LOW, F. E., 1956, *Proc. Roy. Soc.* A, **236**, 112.
COLGATE, S. A., 1958 a, *Geneva*, **32**, 123 ; 1958 b, *Geneva*, **31**, 389.
COLGATE, S. A., FERGUSON, J. P., and FURTH, H. P., 1958, *Geneva*, **32**, 129.
COLGATE, S. A., FERGUSON, J. P., FURTH, H. P., and WRIGHT, R. E., 1958, *Geneva*, **32**, 140.
COLGATE, S. A., and FURTH, H. P., 1958, *University of California Radiation Laboratory, Report* 5392.
COLGATE, S. A., and WRIGHT, R. E., 1958, *Geneva*, **32**, 145.

Conner, J. P., Hagerman, D. C., Honsaker, J. L., Karr, H. J., Mize, J. P., Osher, J. E., Phillips, J. A., and Stovall, E. J., 1958, *Geneva*, **32**, 297.
Cousins, S. W., and Ware, A. A., 1951, *Proc. Phys. Soc.* B, **64**, 159.
Cowling, T. G., 1957, *Magnetohydrodynamics* (New York: Interscience).
Craston, J. L., Hancox, R., Robson, A. E., Kaufman, S., Miles, H. T., Ware, A. A., and Wesson, J. A., 1958, *Geneva*, **32**, 414.
Curran, S. C., Allen, K. W., Bodin, H. A. B., Fitch, R. A., Peacock, N. J., and Reynolds, J. A., 1958, *Geneva*, **31**, 365.
Dattner, A., 1959, *Uppsala*, **2**, 1151.
Dattner, A., Lehnert, B., and Lundquist, S., 1958, *Geneva*, **31**, 325.
Dolgov-Saveliev, G. G., Ivanov, D. P., Mukhovatov, V. S., Razumova, K. A., Strelkov, V. S., Shepelyev, M. N., and Yavlinsky, N. A., 1958, *Geneva*, **32**, 82.
Dreicer, H., 1959, *Phys. Rev.*, **115**, 238 (see also 1960, *Phys. Rev.*, **117**, 329, 343).
Drummond, W. E., and Rosenbluth, M. N., 1960, *Bull. Amer. Phys. Soc.*, **5**, 352.
Ellis, R. A., Goldberg, L. P., and Gorman, J. G., 1959, *Uppsala*, **2**, 1129.
Elmore, W. C., Little, E. M., and Quinn, W. E., 1958, *Geneva*, **32**, 337 (see also 1958, *Phys. Rev. Letters*, **1**, 32).
von Engel, A., 1959, *Nature, Lond.*, **183**, 573.
Etievant, C., Ginot, P., Hubert, P., Rebut, P., Taquet, B., and Torossian, A., 1959, *Uppsala*, **2**, 967.
Fay, H., Hintz, E., and Jordan, H. L., 1959, *Uppsala*, **2**, 1046.
Fetz, H., 1941, *Ann. Phys., Lpz.*, **40**, 579; 1949, *Z. Naturf.*, **4a**, 627.
Finkelstein, D., Sawyer, G. A., and Stratton, T. F., 1958, *Phys. Fluids*, **1**, 188.
Gabriel, A. H., 1959, *Uppsala*, **2**, 829.
Gerjuoy, E., and Stewart, G. W., 1960, *General Atomic Report* GA-1433 (see 1960, *Phys. Fluids*, **3**, 1008).
Gibson, A., 1957, *Venice*, 365.
Gibson, A., Mason, D. W., and Barsanti, G., 1960, *Bull. Amer. Phys. Soc.*, **5**, 342.
Golovin, I. N., Ivanov, D. P., Kirillov, V. D., Petrov, D. P., Razumova, K. A., and Yavlinsky, N. A., 1958, *Geneva*, **32**, 72.
Granovskii, V. L., and Timofeeva, G. G., 1955, *J. Exp. Theor. Phys.*, **28**, 378; 1956, *J. Exp. Theor. Phys.*, **30**, 477.
Griem, H. R., Kolb, A. C., and Faust, W. R., 1959, *Phys. Rev. Letters*, **2**, 281.
Hagerman, D. C., and Mather, J. W., 1958, *Nature, Lond.*, **181**, 226.
Hain, K., Lust, R., and Schluter, A., 1957, *Z. Naturf.*, **12a**, 833.
Haines, M. G., 1959, *Proc. Phys. Soc.*, **74**, 576.
Harding, G. N., Dellis, A. N., Gibson, A., Jones, B., Lees, D. J., McWhirter, R. W. P., Ramsden, S. A., and Ward, S., 1958, *Geneva*, **32**, 365.
Harrison, E. R., 1958, *Phil. Mag.*, **3**, 1318.
Hartman, C. W., Colgate, S. A., and Munger, R., 1960, *Bull. Amer. Phys. Soc.*, **5**, 349.
Hemmings, R. F., Miles, H. T., and Ware, A. A., 1957, *Venice*, 464.
Herdan, R., and Hughes, T. P., 1961, *Proc. Phys. Soc.*, **77**, 987.
Honsaker, J., Karr, H., Osher, J., Phillips, J. A., and Tuck, J. L., 1958, *Nature, Lond.*, **181**, 231.
Hubbard, J., 1958, *Atomic Energy Research Establishment, Report* T/R 2668.
Hughes, T. P., and Kaufman, S., 1959, *Nature, Lond.*, **183**, 7.
Hunt, S. E., 1960, *Proc. Phys. Soc.*, **75**, 596.
Jahoda, F. C., Little, E. M., Quinn, W. E., Sawyer, G. A., and Stratton, T. F., 1960, *Phys. Rev.*, **119**, 843.
Johnson, J. L., 1960, *Bull. Amer. Phys. Soc.*, **5**, 352.
Johnson, J. L., Oberman, C. R., Kulsrud, R. M., and Frieman, E. A., 1958, *Geneva*, **31**, 198.
Jones, W. M., Barnard, A. C. L., Hunt, S. E., and Chick, D. R., 1958, *Nature, Lond.*, **182**, 216.
Jukes, J. D., 1958, *Atomic Energy Research Establishment, Report* GP/R 2293.
Kadomtsev, B. B., 1959, *J. Exp. Theor. Phys.*, **37**, 1096.
Kadomtsev, B. B., and Braginsky, S. I., 1958, *Geneva*, **32**, 233.

KARR, H. J., 1960, *Bull. Amer. Phys. Soc.*, **5**, 349.
KAUFMAN, A. S., HUGHES, T. P., and WILLIAMS, R. V., 1960, *Proc. Phys. Soc.*, **76**, 17.
KAUFMAN, S., and WILLIAMS, R. V., 1958, *Nature, Lond.*, **182**, 557; 1960, *Proc. Phys. Soc.*, **75**, 329.
KERST, D. W., 1958, *Geneva*, **32**, 106.
KIRILLOV, V. D., 1959, *J. Exp. Theor. Phys.*, **37**, 1142.
KNORR, G., 1958, *Z. Naturf.*, **13a**, 941.
KOGAN, V. I., 1958, *Atomnaya Energiya*, **4**, 178; 1959, *Dokl. Akad. Nauk*, **128**, 702.
KOLB, A. C., 1957, *Phys. Rev.*, **107**, 345; 1958, *Geneva*, **31**, 328.
KOLB, A. C., DOBBIE, C. B., and GRIEM, H. R., 1959, *Phys. Rev. Letters*, **3**, 5.
KOLB, A. C., GRIEM, H. R., and FAUST, W. R., 1959, *Uppsala*, **2**, 1037.
KOMELKOV, V. S., 1958, *J. Exp. Theor. Phys.*, **35**, 16.
KOMELKOV, V. S., SKVORTSOV, U. V., and TSEREVITINOV, S. S., 1958, *Geneva*, **31**, 374 (see also KOMELKOV, V. S., 1958, *J. Exp. Theor. Phys.*, **35**, 16).
KRUSKAL, M. D., JOHNSON, J. L., GOTTLIEB, M. B., and GOLDMAN, L. M., 1958, *Geneva*, **32**, 217.
KRUSKAL, M. D., and KULSRUD, R. M., 1958, *Geneva*, **31**, 213.
KRUSKAL, M. D., and OBERMAN, C. R., 1958, *Geneva*, **31**, 137.
KRUSKAL, M. D., and SCHWARZSCHILD, M., 1954, *Proc. Roy. Soc.* A, **223**, 348.
KRUSKAL, M. D., and TUCK, J. L., 1958, *Proc. Roy. Soc.* A, **245**, 222.
KULIKOVSKII, A. G., 1957, *Dokl. Akad. Nauk*, **114**, 984.
KURCHATOV, I. V., 1956, *Atomnaya Energiya*, **1**, 65.
KVARTZHAVA, I. PH., KERVALIDZE, K. N., and GVALADZE, J. S., 1959, *Uppsala*, **2**, 876.
LAING, E. W., 1958, *Atomic Research Energy Establishment, Report* T/M 61 (London: H.M.S.O.).
LAING, E. W., ROBERTS, S. J., and WHIPPLE, R. T. P., 1959, *J. Nucl. Energy*, part C, **1**, 49.
LANDSHOFF, R., 1949, *Phys. Rev.*, **76**, 904.
LATHAM, R., CUZON, F. L., and NATION, J. A., 1959, *Uppsala*, **2**, 884.
LAWSON, J. D., 1957, *Proc. Phys. Soc.* B, **70**, 6.
LEES, D. J., and RUSBRIDGE, M. G., 1959, *Uppsala*, **2**, 954.
LEES, D. J., RUSBRIDGE, M. G., SAUNDERS, P. A. H., and WITALIS, E. A., 1960, *Bull. Amer. Phys. Soc.*, **5**, 381.
LEONTOVICH, M. A., and OSOVETS, S. M., 1956, *Atomnaya Energiya*, **1**, 81.
LILEY, B. S., 1959, *Proc. Instn Elect. Engrs* A, *Suppl.* 2, **106**, 158.
LONGMIRE, C. L., 1958, *Geneva*, **31**, 84.
LUKYANOV, S. YU., and SINITSYN, V. I., 1956, *Atomnaya Energiya*, **1**, 88; 1958, *Geneva*, **32**, 358; *J. Exp. Theor. Phys.*, **34**, 849; 1959, *J. Exp. Theor. Phys.*, **36**, 1621.
LUNDQUIST, S., 1951, *Phys. Rev.*, **83**, 307.
MCWHIRTER, R. W. P., 1959, *Atomic Energy Research Establishment, Report* R-2980.
MCWHIRTER, R. W. P., GRIFFIN, W. G., and JONES, T. J. L., 1959, *Uppsala*, **2**, 833.
MAMYRIN, B. A., 1953, *J. Tech. Phys., Moscow*, **55**, 1914.
MARSHALL, J., 1958, *Geneva*, **31**, 341; 1960, *Phys. Fluids*, **3**, 134.
MATHER, J. W., and WILLIAMS, A. H., 1958, *Geneva*, **32**, 26.
MOROZOV, A. I., 1957, *J. Exp. Theor. Phys.*, **32**, 305.
NAGLE, D. E., QUINN, W. E., RIESENFELD, W. B., and LELAND, W., 1959, *Phys. Rev. Letters*, **3**, 318.
OSOVETS, S. M., NASEDKIN, Y. F., PAVLOV, E. I., PETROV, Y. F., and SCHEDRIN, N. I., 1958, *Geneva*, **32**, 311.
PAYNE, R. M., and KAUFMAN, S., 1959, *Proc. Instn Elect. Engrs* A, *Suppl.* 2, **106**, 36.
PEASE, R. S., 1957, *Proc. Phys. Soc.* B, **70**, 11.
PHILLIPS, N. J., 1959, *Uppsala*, **2**, 849.
PODGORNYI, I. M., KOVALSKII, N. G., and PALCHIKOV, V. E., 1958, *Dokl. Akad. Nauk*, **123**, 825 (see also PODGORNYI, I. M., and CHUVATIN, S. A., 1957, *Dokl. Akad. Nauk*, **117**, 795; KOVALSKII, N. G., PODGORNYI, I. M., and KHVASHCHEVSKII, S., 1958, *J. Exp. Theor. Phys.*, **35**, 940).
POST, R. F., 1956, *Rev. Mod. Phys.*, **28**, 338; 1958, *Geneva*, **32**, 245.

QUINN, W. E., LITTLE, E. M., and RIBE, F. L., 1960, *Bull. Amer. Phys. Soc.*, **5**, 341.
REAGAN, D., 1959, *Nature, Lond.*, **183**, 102 (see also 1959, *Phys. Rev. Letters*, **2**, 82); 1960, *Phys. Fluids*, **3**, 33.
RIBE, F. L., 1959, *Uppsala*, **2**, 1032.
RIDAKOV, L. I., and SAGDEYEV, R. S., 1959, *J. Exp. Theor. Phys.*, **37**, 1337.
ROBERTS, S. J., and TAYLER, R. J., 1957, *Atomic Energy Research Establishment, Report* TR/2138.
ROSE, B., TAYLOR, A. E., and WOOD, E., 1958, *Nature, Lond.*, **181**, 1630.
ROSENBLUTH, M. N., 1956, *Los Alamos Report* LA-2030 (see also 1957, *Venice*, 903); 1958, *Geneva*, **31**, 85.
ROSENBLUTH, M. N., GARWIN, R., and ROSENBLUTH, A., 1954, *Los Alamos Report* LA-1850 (see also 1957, *Magnetohydrodynamics*, edited by R. K. M. LANDSHOFF (Stanford: University Press)).
ROSENBLUTH, M. N., and LONGMIRE, C. L., 1957, *Ann. Phys., N.Y.*, **1**, 120.
ROSENBLUTH, M. N., and ROSTOKER, N., 1958, *Geneva*, **31**, 144.
ROTENBERG, A., LUST, R., RICHTMYER, R. D., LEVY, D., and SUYDAM, B. R., 1960, *Bull. Amer. Phys. Soc.*, **5**, 307.
SAGDEYEV, R. S., KADOMTSEV, B. B., RIDAKOV, L. I., and VEDYONOV, A. A., 1958, *Geneva*, **31**, 151.
SCHLUTER, A., 1950, *Z. Naturf.*, **5a**, 72; 1956, *Proc. Int. Astronomical Union Symposium No.* 6 (Cambridge: University Press), p. 71.
SCOTT, F. R., BASMAN, W. P., LITTLE, E. M., and THOMSON, D. B., 1958, *The Plasma in a Magnetic Field*, edited by R. K. M. LANDSHOFF (Stanford: University Press) (see also SCOTT, F. R., and WENZEL, R. F., 1959, *Phys. Fluids*, **2**, 609).
SHAFRANOV, V. D., 1956, *Atomnaya Energiya*, **1**, 38; 1957, *J. Exp. Theor. Phys.*, **33**, 710 (see also 1957, *Venice*, 990); 1959, *J. Exp. Theor. Phys.*, **37**, 1088.
SIEGBAHN, K., and OHLIN, P., 1958, *Geneva*, **32**, 113.
SIMON, A., 1959, *An Introduction to Thermonuclear Research* (New York: Pergamon Press).
SMITH, C. G., 1941, *Phys. Rev.*, **59**, 997; 1947, *Phys. Rev.*, **71**, 135.
SPITZER, L., 1956, *Physics of Fully Ionized Gases* (New York: Interscience); 1958, *Geneva*, **32**, 181.
STEENBECK, M., and HOFFMANN, K., 1943, *Siemens Technical Report* HW/P1, No. 27.
SUNDSTROM, T., and SVENNERSTEDT, S., 1959, *Nucl. Instrum. Methods*, **4**, 213.
SUYDAM, B. R., 1958, *Geneva*, **31**, 157.
SUYDAM, B. R., LUST, R., RICHTMYER, R. D., and ROTENBERG, A., 1960, *Bull. Amer. Phys. Soc.*, **5**, 307.
TAYLER, R. J., 1957 a, *Proc. Phys. Soc.* B, **70**, 31; 1957 b, *Proc. Phys. Soc.* B, **70**, 1049 (see also 1957, *Venice*, 1067); 1959 a, *Atomic Energy Research Establishment, Report* TR-2787 (London: H.M.S.O.), to be published; 1959 b, *Atomic Energy Research Establishment, Report* L-103, to be published; 1959 c, *Atomic Energy Research Establishment, Report* TR-2786, to be published.
THOMPSON, W. B., 1957, *Proc. Phys. Soc.* B, **70**, 1.
THONEMANN, P. C., 1958, *Geneva*, **31**, 34.
THONEMANN, P. C., BUTT, E. P., CARRUTHERS, R., DELLIS, A. N., FRY, D. W., GIBSON, A., HARDING, G. N., LEES, D. J., MCWHIRTER, R. W. P., PEASE, R. S., RAMSDEN, S. A., and WARD, S., 1958, *Nature, Lond.*, **181**, 217.
THONEMANN, P. C., and COWHIG, W. T., 1951, *Proc. Phys. Soc.* B, **64**, 345.
TONKS, L., 1939, *Phys. Rev.*, **56**, 360.
TUCK, J. L., 1958, *Geneva*, **32**, 3; 1959, *Uppsala*, **2**, 920.
WARE, A. A., 1951, *Phil. Trans. Roy. Soc.*, **243**, 197; 1957, *Engineering*, **184**, 610; 1959 a, *Nature, Lond.*, **183**, 8; 1959 b, *Proc. Instn Elect. Engrs* A, *Suppl.* 2, **106**, 30; 1959 c, *Uppsala*, **2**, 931; 1960, *Bull. Amer. Phys. Soc.*, **5**, 352 (see also 1961, *J. Nucl. Energy*, part C, to be published); 1961, *General Atomic Report* GA-1937.
WARE, A. A., and WESSON, J. A., 1961, *Proc. Phys. Soc.*, **77**, 801.
WATSON, K. M., 1956, *Phys. Rev.*, **102**, 12 (see also BREUCKNER, K. A., and WATSON, K. M., 1956, *Phys. Rev.*, **102**, 19).

COSMIC RADIO WAVES AND THEIR INTERPRETATION

By J. L. PAWSEY and E. R. HILL

Radiophysics Laboratory, Commonwealth Scientific and Industrial Research Organization, Sydney, Australia

CONTENTS

	PAGE
§ 1. Introduction	70
§ 2. General nature of the observations and their interpretation	71
2.1. Basic instrumental limitations	75
§ 3. Theory of the known mechanisms of emission	76
3.1. Thermal emission from fully ionized hydrogen	76
3.2. Synchrotron radiation	79
3.3. The 21-cm line of atomic hydrogen	81
§ 4. Studies of individual galactic objects	82
§ 5. Large-scale structure in our galaxy	84
5.1. Differential galactic rotation and the mass of the galaxy	85
5.2. Spiral and other structure in the galactic disk	87
5.2.1. Spiral structure	87
5.2.2. The shape of the galactic disk	90
5.2.3. The distribution of interstellar hydrogen in the galaxy	92
5.3. Expansion and other phenomena near the galactic centre	93
5.4. The galactic corona	96
§ 6. Comparative observations of normal galaxies	99
6.1. Continuum observations	99
6.2. Hydrogen line observations	102
§ 7. Radio galaxies and cosmology	104
7.1. The nature of radio galaxies	105
7.2. Number counts, background brightness and cosmology	110
Acknowledgments	113
References	113

Abstract. This paper attempts to present the main contributions of radio astronomy to our understanding of our own and external galaxies. Cosmic radio waves originate in interstellar gas by three distinct mechanisms; line emission (from hydrogen), thermal emission by free–free electron transitions, and a non-thermal process believed to be synchrotron emission. Each mechanism is discussed briefly.

Observations, particularly of the line emission, have provided a great extension of knowledge of the distribution and motion of interstellar gas in galaxies. This includes data on rotation leading to estimates of mass distribution, on spiral structure, on the shape of the galactic disk, and on the relative proportions of gas and stars in different parts of our galaxy and in different galaxies.

The synchrotron emission hypothesis implies the existence, in places where non-thermal cosmic radio waves originate, of interstellar magnetic fields and high energy electrons. Two cases of special interest are the remnants of supernovae, which are strong radio sources and may be the primary sources of cosmic rays, and galactic coronas, vast emitting regions which have been found to extend far beyond the stars in our own and some external galaxies.

The discovery of radio galaxies has provided a fresh approach to cosmology. They appear to be observable far beyond current optical limits and, although no definite cosmological information has yet been established from radio evidence, it is probable that significant results may be obtained in the not distant future.

§ 1. Introduction

Just 30 years ago Karl Jansky of the Bell Telephone Laboratories began a series of observations which led to the first recognition of radio waves from beyond the earth and to the subsequent growth of the science of radio astronomy. Several years later, in 1935, Jansky was able to show that the sources of the radio waves he was observing lay in the Milky Way, and he concluded they must lie " either in the stars themselves or in the interstellar matter distributed throughout the Milky Way ".

Radio astronomy is a flourishing science today because it is contributing generously to our physical understanding of very many problems in astronomy. These range over the whole gamut of astronomy, from meteors, through the sun, the planets, our galaxy, and beyond to external galaxies believed to lie further than the limits of all other means of observation. In this paper we propose to restrict ourselves to ' cosmic ' radio waves, i.e. those originating beyond the solar system, and to discuss what we consider (at the time of writing, August 1960) have been the main contributions of radio astronomy to our understanding of the more distant parts of the universe. Some of these contributions involve detailed studies of what was already believed to exist, e.g. spiral structure in our galaxy. Others take the form of completely new discoveries, which, once their nature is understood, may provide information in quite unexpected fields, e.g. supernova radio sources with their bearing on cosmic rays.

Vol. 24
1961

Following Jansky's work the next decisive discovery was that of bright sources of small angular size, ' discrete radio sources ', through the work of Hey, Parsons and Phillips (1946) and of Bolton and Stanley (1948). The small angular size suggested a stellar origin but the very first identifications, of the Crab nebula and the external galaxies NGC 4486 and 5128 by Bolton, Stanley and Slee (1949), were of nebulae, not stars. Today we believe that the great bulk of the cosmic radio waves, both from the discrete sources and the background, originates in interstellar gas. The stars must contribute a quota, the evidence being the observed emission from our local star, the sun, but it appears to be an insignificant fraction of the total. Thus Jansky's second suggestion is now generally accepted.

The current picture of the origin of cosmic radio waves includes three distinct emission processes : (1) a line emission (21 cm) from interstellar atomic hydrogen, and a continuum emission composed of (2) a thermal component from hot fully ionized interstellar gas, and (3) a non-thermal component (very intense at the lower frequencies) due to ' synchrotron ' emission by relativistic electrons spiralling around lines of magnetic force in interstellar space. It follows that radio astronomy contributes directly to our knowledge of physical phenomena in interstellar matter rather than in stars. This circumstance renders radio observations beautifully complementary to optical observations, in which the stars tend to occupy first place. It also gives a bias to the selection of problems in which radio astronomers are interested, a bias towards studies of the interstellar medium and nebulae, and towards the general structure, dynamics, and evolution of galaxies.

The main part of this paper will consist of discussion of particular problems, but in order to establish perspectives we shall begin with several sections of a more general character. We first give a broad outline of the phenomena observed and

an explanation of the main features, with some comments on instrumental capabilities and limitations. We follow this with discussions of the known mechanisms involved in the production of cosmic radio waves, and then proceed to the detailed studies through which we are gaining, or hope to gain, a better physical understanding of our universe.

§ 2. General Nature of the Observations and their Interpretation

Radio telescopes provide measurements of the way in which the strength of cosmic radio waves varies over the sky. The quantities used to specify the strength at any frequency are the brightness (surface brightness) b, which is applicable to extended sources, and the flux density S, which measures the flux from the whole source. The flux density $S = \int_\Omega b d\Omega$ integrated over the solid angle Ω of the source ; m.k.s. units are usual, watt $m^{-2}(c/s)^{-1}$ sterad^{-1} for b, and watt $m^{-2}(c/s)^{-1}$ for S. Alternatively, brightness may be expressed in terms of brightness temperature T_b, the temperature of a black body which would yield the observed brightness. Then

$$b = \frac{2kT_b}{\lambda^2} = 2{\cdot}77 \times 10^{-23} \frac{T_b}{\lambda^2} \text{ m.k.s. units,}$$

where k is Boltzmann's constant and T_b is in degrees Kelvin. These relations are discussed in textbooks, e.g., Pawsey and Bracewell (1955).

The known cosmic radio spectrum consists of the 21-cm hydrogen line and a continuum. The hydrogen line observations show a broad band of emission running around the galactic equator, which is clearly due to the interstellar hydrogen concentrated in the galactic disk, and less intense emission from other directions. The latter must include emission from the near parts of the disk but may include contributions from hydrogen distributed throughout the galactic halo. The emission from near the galactic equator has been extensively studied and has led to knowledge of the general distribution of hydrogen in the disk ; that from other parts of the sky is still essentially a virgin field for research. Twenty-one centimetre emission has also been observed from a few of the nearer external galaxies.

Absorption may be observed if the brightness temperature of incoming radiation incident on the far side of a body of gas is greater than the effective temperature of the gas. The temperature of the interstellar hydrogen is in the vicinity of 100°K, and there are numerous bright discrete sources observed in the continuum for which this criterion is satisfied and which show 21-cm absorption effects.

The continuum extends smoothly over the whole frequency range accessible to observation within the limits imposed by the terrestrial atmosphere. The lower limit, imposed by the ionosphere, is about 1 Mc/s ($\lambda = 300$ m) and the upper, imposed jointly by the weakness of cosmic radio waves and far infra-red atmospheric absorption bands, is about 30 000 Mc/s ($\lambda = 1$ cm). However, the more detailed observations are restricted to the middle of this range. The distribution of the continuum over the sky shows a general background together with numerous small bright regions, the discrete sources.

The background shows a general distribution over the whole sky with a very broad maximum in the direction of the galactic centre, together with a ridge of

intense emission around the galactic equator also showing a marked maximum in the direction of the centre. These features are clearly seen in the very low resolution map of the sky at a frequency of 200 Mc/s shown plotted in galactic coordinates† in Fig. 1. The obvious relationship to the disposition of the galaxy clearly substantiates Jansky's conclusion that the bulk of the emission must originate in our galaxy. Comparison with the known distribution of stars shows that the sources of the background are distributed differently from the stars, and it is now established that they are concentrated in at least two sub-systems: a relatively strongly emitting region in the galactic disk, and a roughly spherical 'corona' extending to distances far beyond the bulk of the stars. The resolution of the

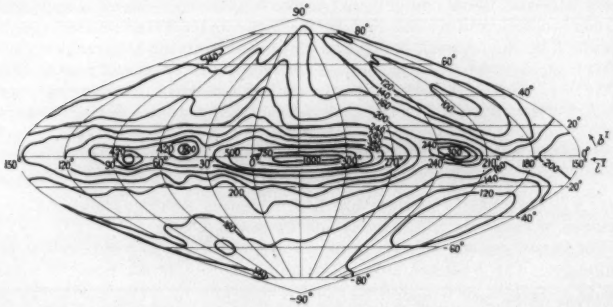


Fig. 1. A low resolution map of brightness temperature (°K) over the sky at 200 Mc/s compiled by Dröge and Priester (1956). The galactic centre lies near the middle of the map at $l^{I} = 327°{\cdot}7$, $b^{I} = -1°{\cdot}4$.

equipment used in deriving the map of Fig. 1 was inadequate to show the narrowness of the disk or the considerable degree of irregularity which exists in both corona and disk. Of the gross features which are visible, two of the known maxima along the equator, those at $l^{I} = 232°$ and 48°, are probably in directions in which we look along spiral arms; the origin of the spur running north from the equatorial ridge at about $l^{I} = 0$ is the subject of controversy (see discussion, Tunmer 1959).

The integrated radiation from the whole sky, which is dominated by the corona, shows a spectrum of the form $T_b \propto f^{-\alpha}$, where T_b is the brightness temperature at frequency f and α is about 2·5. At the lower frequencies, values of T_b are very high, e.g. 10^5 to 10^6 deg K at 20 Mc/s. The high brightness temperature, and the fact that α exceeds 2, are incompatible with an origin in thermal emission from

† Astronomy is in a stage of transition between an old and a revised system of galactic coordinates. In 1958 the International Astronomical Union decided to adopt a revised system (Blaauw, Gum, Pawsey and Westerhout 1959) because in the previous (Lund) system the coordinate pole differed from the true galactic pole by about 1°·5. In the new system this deviation is removed and the zero of longitude, which was previously in an arbitrary position, is changed to make it pass through the galactic centre. The revised system is much more convenient for galactic studies and we should have preferred to use it exclusively. However, existing maps are all in the old system and we have compromised by using both systems. In the old system, in which longitude and latitude are designated l^{I} and b^{I}, the galactic centre is in the direction (327°·7, −1°·4); in the revised system (l^{II}, b^{II}), the centre is at (0°; 0°).

ionized interstellar gas (see § 3.1), the mechanism originally suggested by Reber (1940). The bulk of the emission must originate in a non-thermal process and, as mentioned above, the ' synchrotron ' mechanism is now considered most plausible.

The pattern of radiation from the galactic disk is different at different frequencies, suggesting the operation of more than one factor. This complication is believed due to the presence of widely distributed non-thermal sources similar to those in the corona, together with clouds of ionized interstellar gas (HII regions), which are known to be common in the galactic disk. The absorption coefficient of fully ionized gas varies as f^{-2} and quantitative considerations (see § 3.1) indicate that at low frequencies, e.g. 20 Mc/s, the optical depth of a typical HII region may reach values greater than unity. Hence such a region, when seen against a non-thermal background of brightness temperature in excess of the temperature of the HII region, normally about 10 000°K, will appear as an absorbing, or even opaque, object. Figs 2(*a*), (*b*) and (*c*) show radio maps of a part of the Milky Way obtained at frequencies of 1390, 85 and 19 Mc/s respectively with radio telescopes each having a resolution of about 1 degree. The maps each show a highly irregular structure but the striking feature is that, whereas the two higher frequencies show a ridge along the Milky Way, the lowest shows an initial rise towards the ridge but instead of the ridge there appears an irregular trough, or series of depressions. This trough is due to absorption in the HII regions concentrated near the galactic plane.

The discrete sources are distributed over the sky, both within and outside the Milky Way. They range in diameter from degrees down to the instrumental limit of about 10 seconds of arc. But all the intense ones have been resolved ; they are radio nebulae, not radio stars. In 1952 Mills showed that, statistically, sources fall into two distinct classes : those near the Milky Way, which he termed Class I sources, and those remote from it which he termed Class II. It is now clear that most, or perhaps all, Class II sources are external galaxies ; Class I sources are mainly galactic objects but include a few exceptional external galaxies which happen to be near the galactic equator, e.g. the Cygnus A source. The map in Fig. 1 is of such low resolution that only one discrete source, Cassiopaeia A (at $I^{\text{II}} \simeq 80°$), the strongest of all, stands out above the other features. In the high resolution maps of Fig. 2 many discrete sources are prominent.

Among the galactic sources two classes are known : non-thermal galactic sources and HII regions. The latter emit thermally and are recognized by their having spectra of the expected theoretical form ; some coincide in position with optically visible HII regions. Most, or perhaps all, of the remaining sources have a spectrum similar to the galactic corona, and therefore must be non-thermal.

Because the high frequency part of the spectrum of an HII region is of the form $T_b \propto f^{-2}$ (see § 3.1), whereas for the non-thermal sources it is $T_b \propto f^{-2\cdot7}$ approximately, the HII regions tend to become relatively more intense at high frequencies. In practice at low frequencies, e.g. 20 Mc/s, HII regions appear generally in absorption against the background ; at intermediate frequencies, e.g. 100 Mc/s, they tend to merge into the background and nearly all observed sources are non-thermal. At high frequencies, e.g. 1400 Mc/s, HII regions are relatively much more intense with respect to both background and non-thermal sources, and roughly equal numbers of thermal and non-thermal sources are observed in surveys of the Milky Way.

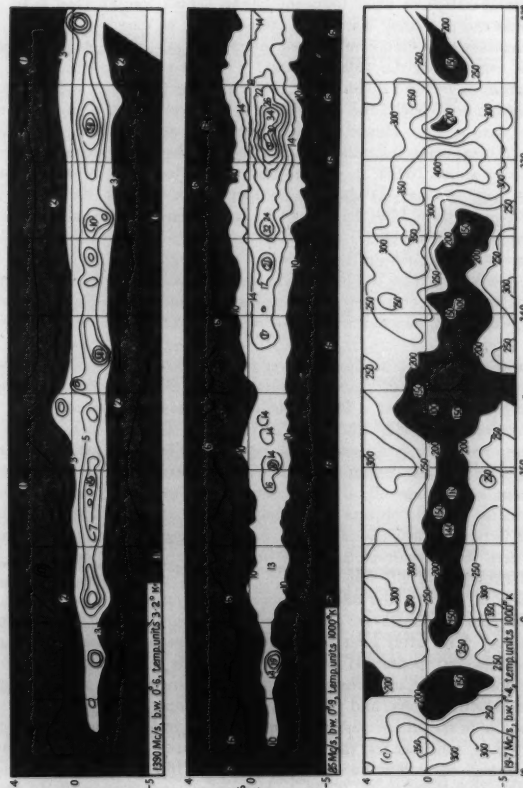


Fig. 2. Simplified maps of the same section of the Milky Way at widely different frequencies, (*a*) from Westerhout (1958), (*b*) from Hill, Slee and Mills (1958), and (*c*) from an unpublished provisional map drafted by the late C. A. Shain.

It is a striking fact that, excluding the sources we believe to be HII regions, both galactic and extra-galactic discrete sources should have spectra of *approximately* the same form and that this form should approximate also to that shown by the galactic background. The agreement is not exact but is sufficient to point strongly to a common mechanism for all three : galactic background, galactic sources and extra-galactic sources. As noted above it is supposed that the common mechanism is the synchrotron mechanism. The power law is then accounted for as due to a power law spectrum of energies among electrons (similar to that observed for cosmic rays) and differences in exponent as due to differences in the electron-energy spectrum in the various regions of origin. The solar spectrum is quite different.

2.1. *Basic Instrumental Limitations*

Although a detailed discussion of radio telescopes is inappropriate here, some reference to their capabilities and limitations is desirable.† The basic characteristics of a radio telescope are its angular resolution and sensitivity. For survey studies of the background and sources, resolutions (measured by beam width) of about half a degree are now available from arrays and paraboloids at high frequencies, and from ' enhanced resolution ' systems such as the Sydney ' Mills Cross ' and the Cambridge ' aperture synthesis ' techniques at low frequencies. An order of magnitude improvement in resolution at both high and low frequencies can be expected in the next few years. In the study of discrete sources considerably higher resolution has been obtained using two-aerial interferometers which have provided precise positions and detailed brightness distributions of the stronger sources. Extension of this technique to fainter sources is restricted by blending effects ; for these only crude angular sizes down to 10 seconds of arc have been measured.

The signal conveyed by an aerial to its receiver is proportional to T_b when the source is more extensive than the beam width, and to AS, where A is the effective area of the aerial and S the source flux density, when the source is much smaller than the beam width. To be detected by the receiver these powers must adequately exceed the inherent fluctuations given by $T/\sqrt{(B\tau)}$ where T is a measure of the total noise, internal plus external, and B and τ are the receiver bandwidth (c/s) and time constant (sec) ; $1/\sqrt{(B\tau)}$ is of the order of 10^{-3}. Thus $T/\sqrt{B}$ is a figure of merit for a receiver for continuum studies (B limited by receiver), and T for H-line studies (B chosen to suit profile). The current situation with regard to T is indicated in Fig. 3, where receiver noise temperatures for conventional receivers are compared with the external cosmic noise. Below about 200 Mc/s sensitivity is irrevocably limited by cosmic noise. Above 200 Mc/s it is limited by receiver noise and improvement is possible. A further technical limitation, important at the higher frequencies, is the difficulty of measuring absolute, as opposed to relative, brightness temperatures to better than a few degrees. For such reasons continuum observations at the higher frequencies have been restricted to the bright regions in the Milky Way and discrete sources and have been expressed as temperature excesses over colder parts of the sky.

† For further details regarding aerials, see Christiansen (1959), for receivers see Pawsey and Gardner (1959).

Recently, greatly improved receivers, masers and parametric amplifiers, have been invented which promise one or two orders of magnitude improvement in sensitivity. Some sample performance figures are shown in Fig. 3. Potentially such receivers are capable of giving vastly improved sensitivity for H-line studies and, with the large paraboloids, of opening up the decimetre wavelength range to really fruitful observation. However, they have not yet made their impact on radio astronomy.

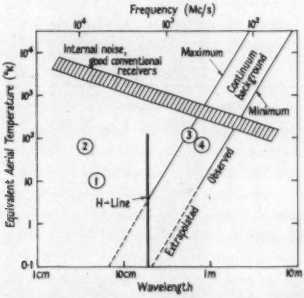


Fig. 3. Levels of cosmic line and background radiation in relation to internal noise generated in conventional receivers. Performances of four new experimental receivers are also shown: (1) T.W. maser, (2) cavity maser, (3) Adler tube, and (4) T.W. crystal diode parametric amplifier.

§ 3. Theory of the Known Mechanisms of Emission

We now digress briefly to outline the theory of the three known mechanisms of emission. The interstellar gas is mainly hydrogen. In many parts it is in a neutral atomic state; such regions are known as HI regions and the hydrogen emits and absorbs the 21-cm line, but no appreciable light. In other parts the gas is almost fully ionized; these regions are known as HII regions and they emit and absorb thermally a radio continuum. They give strong line emission in the optical spectrum and are also called emission nebulae. The theory of these processes is well known, see for example Pawsey and Bracewell (book, 1955), and we shall merely recapitulate the key equations and indicate significant considerations. In the third case, synchrotron radiation, we have included a qualitative description of the physical basis of the mechanism.

3.1. *Thermal Emission from Fully Ionized Hydrogen*

Under conditions applicable in HII regions the dominant thermal emission mechanism in the radio spectrum is that due to the acceleration of electrons as they collide with positive ions (free–free transitions). The absorption coefficient κ is given by

$$\kappa = \frac{\zeta N_e^2}{f^2 T_e^{3/2}},$$

where ζ is a very slowly varying parameter approximating 0·13 c.g.s. units for conditions of interest here, N_e = electron density, f = frequency, and T_e = electron temperature in °K.

The optical depth τ to a distance x_0 along a line of sight is given by

$$\tau = \int_0^{x_0} \kappa \, dx = \frac{\zeta}{f^2} \int_0^{x_0} \frac{N_e^2}{T_e^{3/2}} \, dx,$$

which, for T_e constant, reduces to

$$\tau = \frac{\zeta}{f^2 T_e^{3/2}} \int_0^{x_0} N_e^2 \, dx.$$

This integral appears in optical studies of nebulae where, with N_e expressed in electrons cm^{-3} and distance x' in parsecs, it is called the ' emission measure ' (e.m.). With these units and with $T_e = 10^4$ deg K this equation reduces to

$$\tau = \frac{4{\cdot}1 \times 10^{11}}{f^2} \int_0^{x_0} N_e^2 \, dx' = \frac{4{\cdot}1 \times 10^{11}}{f^2} \times \text{(e.m.)}.$$

To illustrate expected trends it is instructive to consider the values of τ computed for, (1) a typical HII region and, (2) a typical 10 kpc path in the galactic disk. Parameters adopted are those suggested by Oort (1959).

Computed Values of Optical Depth τ

	Optical depth τ		
Frequency f (Mc/s)	20	85	1420
A typical HII region ($N_e = 10$ cm^{-3}, depth 34 pc ; (e.m.) = 3400 cm^{-6} pc)	3·5	0·2	0·0007
10 kpc of galactic disk (39 of above regions kpc^{-3})	1·2	0·07	0·0002

Remembering the wide variation between individual HII regions, it appears that at 20 Mc/s many HII regions will have large optical depths and will be effectively opaque to the intense radiation from beyond. It also appears that visibility of the more distant parts of the galactic disk may be severely restricted. At frequencies of 85 Mc/s and higher, on the other hand, opacity should only be important in exceptional cases.

For an isolated body of gas in the absence of any other source of radiation the brightness temperature T_b due to thermal emission is given by

$$T_b = \int_0^{\tau_0} e^{-\tau} T_e \, d\tau.$$

For the isothermal case this reduces to

$$T_b = T_e\{1 - \exp(-\tau_0)\},$$

where τ_0 is the optical depth to the far side. For $\tau_0 \ll 1$ this reduces to

$$T_b = T_e \tau_0,$$

and for $T_e = 10\,000°\text{K}$ as before,

$$T_b = \frac{4{\cdot}1 \times 10^{15}}{f^2} \times (\text{e.m.}).$$

The form of the variation of T_b with frequency is important in considering mechanisms of emission and is illustrated in Fig. 4(*a*) for the isothermal case. At low frequencies T_b is asymptotic to T_e; at high frequencies the limiting slope is given by f^{-2}. Even if the gas is not isothermal the slope of the spectrum can never become steeper than this limit. Figure 4(*a*) also illustrates the typical spectrum of a non-thermal source: slope steeper than f^{-2}, and T_b values exceeding 10 000°K.

If brightness is expressed, not in temperature but in m.k.s units, an additional f^2 term is introduced since $b = 2kT_b/\lambda^2$. Thus for low frequencies where $\tau \gg 1$,

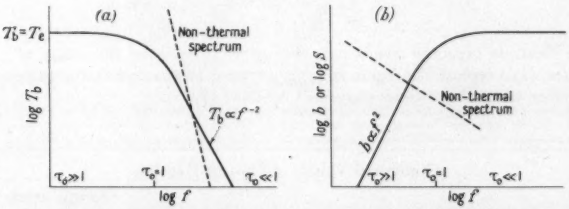


Fig. 4. The variation of (*a*) T_b, and (*b*) b and S, with frequency for thermal emission from an isothermal cloud of fully ionized hydrogen. A typical non-thermal spectrum is also shown.

b is proportional to f^2, and at high frequencies where $\tau \ll 1$, b approaches a constant value. The flux density S, being the integrated value of b over the solid angle subtended by the source, shows a similar type of variation. The behaviour of b and S is shown in Fig. 4(*b*). The detailed form of these curves depends on details of the density distribution in the source since different parts reach saturation at different frequencies.

The above discussion refers to an isolated nebula. However, in practice a complication generally arises due to the nebula being embedded in a non-thermally emitting medium. Considering a nebula in front of a background whose temperature is T_b, then what is measured is the apparent excess temperature ΔT of the cloud over the adjacent background, i.e.

$$\Delta T = T_e\{1 - \exp(-\tau_0)\} + T_b \exp(-\tau_0) - T_b = (T_e - T_b)\{1 - \exp(-\tau_0)\}.$$

A similar expression holds for the ' apparent flux density '. Thus a nebula appears in emission, or absorption, or not at all, depending on T_e in relation to T_b as described in § 2. It should be noticed that, at high frequencies, where $T_e \gg T_b$, this result reverts to that given earlier because then the non-thermal emission is weak.

3.2. *Synchrotron Radiation*

This type of emission first became of practical importance in connection with the emission of light, and consequent energy loss, in electron beams in synchrotrons —hence the term ' synchrotron radiation '. The suggestion that this mechanism was important also in the generation of cosmic radio waves was originally made by Alfvén and Herlofson (1950) to explain the nature of discrete radio sources ; subsequently Kiepenheuer (1950) applied the theory to explain the background of radio emission from the whole galaxy. Possibly because these theories appeared before enough was known regarding radio sources or galactic emission, these

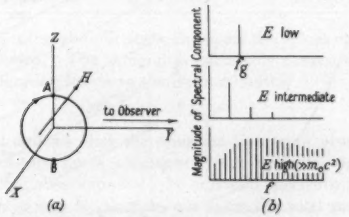


Fig. 5. Illustrating synchrotron radiation.

works received little attention for some years except from a number of theoreticians in the U.S.S.R. These included Ginsburg and Shklovsky, and the latter suggested that not only the radio waves but the light from the Crab nebula should arise in this manner. Subsequent observation of polarization of optical emission from the nebula finally led to the general acceptance of the mechanism.

We shall now discuss this mechanism, i.e. the emission of radiation by relativistic electrons spiralling in a magnetic field. The distinctive factors may be recognized by considering the changes in the nature of the emission in the Y-direction from the electron shown in Fig. 5(a) as its energy E is progressively increased. The electron is supposed to move in a circle in the YZ-plane under the influence of a magnetic field along the X-axis. It radiates because it is accelerated, and at low energies ($\ll m_0 c^2$) the radiated field in the Y-direction is sinusoidal with positive and negative peaks at times corresponding to the times at which the electron passes A and B. As the energy is increased the velocity approaches the speed of light and the important new factor is that emission in the Y-direction from the series of points in the vicinity of A is in phase and so reinforces strongly. In consequence the emission becomes peaked at this point in the cycle, and at very high energies ($\gg m_0 c^2$) degenerates into a series of short pulses centred on the times of passing A. Furthermore, the total emission increases greatly and the gyro-frequency decreases. The corresponding effect on the spectrum is shown in Fig. 5(b). The low energy case gives a single frequency at the gyro-frequency ; the intermediate energy shows a somewhat lower fundamental frequency and a few harmonics ; and the high energy case a large number of harmonics having an envelope of the form discussed later and extending to quite high frequencies. The

emission is plane polarized with the E-vector in the YZ plane, and is concentrated in a narrow cone, of semi-angle of the order of $m_0 c^2/E$ about the direction of motion of the electron. For many purposes, since the harmonics lie so close together, we may consider the radiation as if it were continuous. This approach is adopted below.

The theory of this emission is complicated, and we shall merely quote from Oort and Walraven (1956), and Oort (1959), some results which are of particular

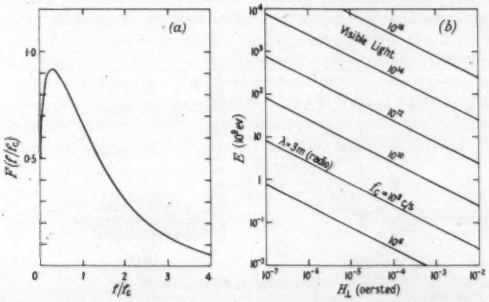


Fig. 6. Synchrotron radiation; (*a*) the spectrum emitted by an individual electron; (*b*) the critical frequency f_c as a function of magnetic field and electron energy.

significance in their simplest and most useful form. The power emitted by a high energy electron in ergs per second per c/s at frequency f is given by

$$P_f = CH_\perp F(f/f_c),$$

where $C = 2{\cdot}343 \times 10^{-22}$, $H_\perp$ is the component of the magnetic field (in oersteds) perpendicular to the direction of motion of the electron, while $F(f/f_c)$ describes the spectrum of the radiation and is shown graphically in Fig. 6(*a*). The ' critical frequency ' f_c is a convenient frequency lying in the emitting range and is defined in c/s by $f_c = LH_\perp(E_{\text{GeV}})^2$, where $L = 1{\cdot}608 \times 10^{13}$, and E_{GeV} is the total energy E of the electron $[= m_0 c^2/(1-v^2/c^2)^{1/2}$, where m_0 is the rest mass, and v the velocity] in units of 10^9 ev. f_c is shown as a function of E_{GeV} and $H_\perp$ in Fig. 6(*b*).

The spectrum is seen to be a broad spectrum extending from frequencies near zero to a small multiple of the critical frequency. Hence f_c, which may be read off Fig. 6(*b*), may be regarded as giving a rough upper limit to the frequency spectrum. A 1 Gev electron, for example, will radiate in the 100 Mc/s range only in magnetic fields of the order of 10^{-5} oersted or greater. Similarly visible light would require, for field strengths of the order of 10^{-5} oersted, electrons with energies of the order of 10^3 Gev.

The emission of energy, integrated over all frequencies, is proportional to $H_\perp^2 E^2$, leading to the decay of electron energy from an initial value E_0 (in units of 10^9 ev) given by

$$E_{\text{GeV}} = E_0/(1 + AH_\perp^2 E_0 t),$$

where E_{GeV} is the energy after time t, and $A = 3{\cdot}793 \times 10^6$. Thus a 1 GeV electron in a field of 10^{-4} oersted loses half its energy in $0{\cdot}835 \times 10^6$ years.

We turn now to an assembly of electrons which we assume to have a distribution of energies within limiting values E_1 and E_2 of the form

$$N(E)\,dE = kE^{-\beta}\,dE \quad \text{for} \quad E_1 < E < E_2,$$

and

$$N(E)\,dE = 0 \quad \text{for} \quad E < E_1 \quad \text{and} \quad E > E_2.$$

The factor k is proportional to the total energy of the electrons. The combined emission is given, according to Woltjer (1958), by

$$J_f = 1{\cdot}17 \times 10^{-22} L^{(\beta-1)/2} k f^{-(\beta-1)/2} H_\perp{}^{(\beta+1)/2} \int_{\alpha_2}^{\alpha_1} \alpha^{(\beta-3)/2} F(\alpha)\,d\alpha,$$

where $\alpha = f/f_c$ and the integration limits α_1 and α_2 are the values with f_c corresponding to E_1 and E_2 respectively. The integral is effectively constant over a wide range of f between the integration limits so that approximately

$$J_f \propto k f^{-(\beta-1)/2} H_\perp{}^{(\beta+1)/2}.$$

From the relation it follows that the emission spectrum gives the form of the electron energy spectrum, i.e. a spectrum $f^{-(\beta-1)/2}$ implies an energy spectrum $E^{-\beta}$. It also shows that observations of the emission spectrum alone give a relation between particle energy and $H_\perp$, but do not allow the total particle energy (fixed by k) to be determined separately from $H_\perp$—this is an important characteristic of synchrotron emission.

If we assume an injection of electrons with random directions into a uniform magnetic field the observed intensity is given by

$$J_f \propto k f^{-(\beta-1)/2} (\sin\theta)^{(\beta+1)/2},$$

where θ is the angle between the direction of observation and the magnetic field. This power law is in accord with commonly observed spectra of cosmic sources, for which the exponent is typically $-0{\cdot}7$ corresponding to $\beta = 2{\cdot}4$. The directional effect predicted is a broad maximum in the directions perpendicular to the field, falling to half intensity at angles somewhat more than 45° from the maximum. This prediction is critically dependent on the assumption of random directions. The radiation should also be substantially plane polarized in the plane perpendicular to the component of the magnetic field transverse to the line of sight.

3.3. *The 21-cm Line of Atomic Hydrogen*

This line is a hyperfine structure line which is associated with a transition between two sub-levels in the ground state. As before we shall limit this discussion to a summary of the facts most often required for astronomical interpretation (see Wild 1952).

The rest frequency of the line is 1420·4056 Mc/s. The natural line width is exceedingly small, 5×10^{-16} c/s, and the observed line widths are determined by other causes of which Doppler effects are paramount. At this frequency a velocity of 1 kilometre per second corresponds to a frequency shift of 4·73 kc/s, and conversely 1 kc/s to 0·211 kilometres per second. Typical galactic profiles are from a hundred to a thousand kc/s in width.

Emission and absorption depends on the relative populations of the two atomic levels concerned. Under equilibrium conditions this ratio is determined by the temperature and it is convenient to describe the state of the gas in terms of the equilibrium temperature T_s which would give the actual ratio. This is termed the 'excitation temperature' (or 'spin temperature'). It is believed that T_s approximates closely to the kinetic temperature of the gas for interstellar conditions (Purcell and Field 1956). Emission and absorption are given quantitatively by the usual thermodynamic relations involving optical depth τ_f and temperature T_s. The absorption coefficient κ_f at frequency f is given by

$$\kappa_f = 8{\cdot}0 \times 10^3 \frac{n_f}{T_s} \text{ parsec}^{-1},$$

where n_f is the number of atoms per cm^3 in the ground state with line-of-sight velocities in the range corresponding to unit frequency band about the frequency f. In the isothermal case, the optical depth to a distance x_0 is given by

$$\tau_f = \int_0^{x_0} \kappa_f \, dx = \frac{8{\cdot}0 \times 10^3}{T_s} \int_0^{x_0} n_f \, dx.$$

The brightness temperature $T_b(f)$ at this frequency due to emission from an isolated isothermal cloud is given in terms of the optical depth through the cloud by

$$T_b(f) = T_s\{1 - \exp(-\tau_f)\}.$$

For $\tau_f \gg 1$ this reduces to $T_b(f) \simeq T_s$, and for $\tau_f \ll 1$ to $T_b(f) = T_s \tau_f$. The latter gives

$$T_b(f) = 8{\cdot}0 \times 10^3 \int_0^{x_0} n_f \, dx.$$

If we consider the whole line profile

$$\int_0^{\infty} T_b(f) \, df = 8{\cdot}0 \times 10^3 \int_0^{x_0} n \, dx,$$

where n is the total number of hydrogen atoms per cm^3 in the ground state along the line-of-sight. Thus the area under the profile is a measure of the total number of hydrogen atoms in a column of unit cross section along the line-of-sight independently of their temperature, provided self-absorption is negligible. Measurement of both T_s and $\int n \, dx$ is possible for a cloud of hydrogen lying between the observer and a bright discrete source of small diameter, because in this case absorption and emission can be separately determined.

§4. Studies of Individual Galactic Objects

The distribution of both 21-cm and continuum radio waves over the sky shows a considerable degree of irregularity, and there is little doubt that very many of these irregularities are associated with individual galactic objects and can be used in their study. Many such studies have been reported, but so far broad physical conclusions have been reached only in the cases of two classes of object: supernova remnants and emission nebulae. The first originated in the identification of the radio source Taurus A as the Crab nebula by Bolton and Stanley (1949), and the second in the detection of a series of emission nebula at 10 cm by Haddock, Mayer

and Sloanaker (1954). We shall restrict detailed discussion to these two classes, but should like to mention studies of hydrogen in stellar associations and galactic clusters (Menon 1958, Drake 1959 a) and continuum observations of planetary nebulae (Drake and Ewen 1958) as interesting lines of development. Novae have been sought but not observed.

Radio emission from emission nebulae is thermal, being that expected from the optical model of a highly ionized gas with the characteristic electron temperature of about 10^4 °K. Radio observations are thus a potential source of information to help fill in optical conclusions regarding physical conditions within emission nebulae. Radio measurements of electron temperature by Mills, Little and Sheridan (1956) and Wade (1958, 1959 a) give values close to those expected. On the other hand radio estimates of electron densities within nebulae are dependent on assumptions concerning density irregularities (see § 5.2.3), and derived masses of nebulae are very uncertain. The study of the Orion nebula (Fig. 7, Plate I) by Osterbrock and Flather (1959) using combined radio and optical data has indicated how the problem of irregularities may be solved and estimates of masses improved. In addition, radio techniques have brought about a substantial increase in the potential range of detection of emission nebulae, since high frequency observations can detect nebulae which would otherwise be completely obscured by intervening dust clouds. Nearly half of the 60 or so nebulae detected by radio means (Westerhout 1958, Mills, Little and Sheridan 1956) were previously unknown optically. The study of these distant invisible nebulae poses serious problems—for instance how are their distances to be found from radio observations alone? Shain (1958) and Mills (1959 a) have suggested possible methods depending on the distribution of non-thermal emission in front of and behind the nebula.

The gigantic stellar explosions known as supernovae involve the ejection of a substantial fraction of the exploding star's mass with velocities of several thousands of kilometres per second, and lead to light emission which for a few weeks is comparable with the entire emission from the galaxy. Supernovae are of two types. Type I appear normally in the halo and belong to Baade's Population II of old stars, while Type II, which eject more massive shells at higher velocities and occur in the disks of spiral galaxies, are Population I objects. There are three optically well documented supernova explosions in our galaxy—the Chinese supernova of 1054 A.D. (whose remnant is the Crab nebula), Tycho's star (1572 A.D.), and Kepler's star (1604 A.D.)—all are Type I supernovae and sources of non-thermal radio emission. As far as we know, no Type II supernova eruptions have ever been observed in our galaxy, but a number of radio sources are now believed to represent remnants of Type II supernovae. These include Cassiopaeia A, the strongest radio source (see Minkowski 1959), and possibly several radio sources associated with faint fragmentary nebulosities such as IC 443 and the Cygnus Loop (Minkowski 1958). Recently, Mills (1959 a, b, c) has given evidence indicating that many of the 30 or so non-thermal radio sources near the galactic equator are remnants of very distant Type II supernovae.

The brightest supernova remnant, optically, is the Crab nebula (Fig. 8, Plate II) and through a study of this object it is hoped will come the key to understanding this class of radio source. The nebula consists of an expanding shell of filaments showing line emission, and an amorphous inner mass with a continuous

spectrum. The most widely accepted explanation of the origin of both the radio and optical emission from the inner mass, due to Shklovsky (1953 a), is synchrotron emission from relativistic electrons and positrons spiralling through a magnetic field in the nebula. The principal evidence supporting this hypothesis has come from the discovery of optical linear polarization by Dombrovsky (1954) and subsequent confirmation by Oort and Walraven (1956) and Woltjer (1957) among others. Woltjer's map of the optical polarization over the nebula is shown in Fig. 9 (Plate IV). More recently linear polarization at radio frequencies above 3000 Mc/s has been detected (Mayer, McCullough and Sloanaker 1957, Kuzmin and Udaltsov 1959, and Mayer and Sloanaker 1959).

The synchrotron emission establishes the form of the energy spectrum of the relativistic particles and gives a relation between their concentration and the magnetic field—the latter can only be individually determined from additional considerations. Shklovsky (1953 a), Oort and Walraven (1956) and Woltjer (1958) all suggest magnetic fields of 10^{-3} to 10^{-4} oersted, which require particle energies of the order of 10^{9} ev for the radio emission and 10^{12} ev for the optical. The most striking result is that for all plausible magnetic fields the total magnetic and particle energies comprise about 10^{48} ergs (comparable with the kinetic energy of the shell), and about 10^{-4} of the total nuclear energy presumed to be available from the complete conversion of hydrogen in the original star into carbon. The means by which so much energy is made available in the form of fast particles and magnetic field is by no means clear. Suggestions for the origin of the energetic particles include production of fast thermal particles on impact of the ejected shell with the interstellar medium (Shklovsky 1954), thermonuclear reactions stimulated by the impact (Greenstein 1957), the products of the decay of radioactive elements formed at the time of explosion (Woltjer 1958), or the result of a primary flux of protons colliding with nuclei forming mesons which later decay into positrons and electrons (Burbidge 1958). In all cases Fermi statistical acceleration is required to bring the particles to optically radiating energies. Some consider variabilities in the nebula, described by Oort and Walraven (1956), to be manifestations of this process. The nature and origin of the magnetic field is another enigma—it is clear that it has some degree of uniformity (see Fig. 9), and also that it cannot be that of the original star. Woltjer (1958) suggests that it is a force-free field controlled by currents in the outer shell of the nebula.

Finally, an interesting by-product of the above theory of supernovae must be mentioned. It is the possibility that the fast particles required to explain the galactic corona (see § 5.4), and the primary cosmic rays, may have originated within supernovae (Shklovsky 1953 b). This hypothesis is still highly speculative but we know of no other more plausible one. The synchrotron mechanism implies that, using cosmic radio waves, we can ' see ' cosmic rays or associated particles throughout our galaxy. Shklovsky (1960) in his book gives a most interesting account of these developments from the point of view of the Russian pioneers.

§ 5. Large-scale Structure in our Galaxy

We turn now from individual objects in the galaxy to the study of its *large-scale structure*, particularly to the distribution and motion of interstellar gas in the galactic

disk, and to the nature of the *galactic corona*, which is one of the fundamental discoveries of radio astronomy. Evidence on the first has been derived mainly from 21-cm line studies, on the latter from the continuum. In these studies radio observations display their characteristic feature—the lack of obscuration by dust clouds permitting a clear view to the farthest limits of the galaxy.

It should be noted that wherever possible in this section we have used the revised galactic coordinate system (see § 2.0) where longitude and latitude are denoted by l^{II} and b^{II} and the galactic centre is in the direction (0°, 0°).

5.1. *Differential Galactic Rotation and the Mass of the Galaxy*

Observations of the 21-cm line at various galactic longitudes show generally a maximum of emission at a latitude near the galactic equator, due to the concentration of hydrogen in the galactic disk. Typical profiles for various points on the galactic equator are shown in Fig. 10.

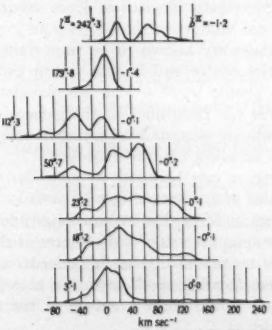


Fig. 10. Typical equatorial galactic H-line profiles (Oort 1959).

The widening of the line is due to Doppler shifts, and the intensity at a particular frequency is a measure, subject to small corrections for saturation, of the number of hydrogen atoms in the appropriate frequency range integrated along the line-of-sight (see § 3.3). Peaks in the profiles therefore indicate concentrations of atoms with corresponding line-of-sight velocities. The velocity with respect to the sun is due to various causes, but fortunately it is dominated by differential galactic rotation. This effect is zero in directions towards the centre and anticentre ($l^{II} = 0°$ and 180°) and the exceptional narrowness of the profile at $l^{II} = 180°$ (Fig. 10) illustrates this point.

Assuming circular orbits in the galactic plane, the line-of-sight velocity V, relative to the sun, of a body of neutral hydrogen at longitude l^{II} and distance R from the galactic centre is given by

$$V = R_8 \sin l^{II}(\omega_R - \omega_8),$$

where ω_R is the angular velocity of rotation about the centre of the galaxy, and R_s and ω_s refer to the sun. This relation may be used to find ω_R from the measured velocity V if R_s and ω_s are known ; if ω_R is a known function of R, R also follows. For $R > R_s$ the position determined by R and l^{II} is unique, but for $R < R_s$ an ambiguity arises corresponding to the two intersections of the circle of radius R with the line-of-sight. Thus, in the absence of motions other than circular orbital motion, and subject to this ambiguity which can be resolved to some extent on other evidence, Doppler shifts may be used to determine distances and hence positions.

In practice such distance determinations are complicated by departure from circularity of the orbits of the sun and gas clouds. Allowance has been made for the peculiar motion of the sun in profiles such as those of Fig. 10, which indicates velocities relative to the local standard of rest, i.e. the average motion of local stars. Random peculiar motions of individual gas clouds give a random scatter of derived positions and so blur, but do not falsify, derived pictures of the general distribution of neutral hydrogen. Systematic deviations from circular orbits, either by the gas clouds or by stars in the vicinity of the sun, would invalidate distance determinations. For instance there are known to be important expansion effects in the region within 3 kpc of the centre and in this region the method is inapplicable. The long wings of the profile at $l^{II} = 3°{\cdot}1$ are associated with these expansion motions (see § 5.3). Over the remainder of the galactic disk, however, the mean orbital motions are tentatively accepted as approximately circular and the above principles are applied in locating neutral hydrogen.

The radio measurements can be used to help derive the rotational model. Between the sun's distance and points about 3 kpc from the centre of the galaxy ω_R increases with decreasing R so that at any longitude between 20° and 90° of the galactic centre the maximum value of V occurs at the ' tangential point ' T, where the line-of-sight is tangential to a circle concentric with the galactic centre. Hence, if hydrogen were distributed all over the plane, the resulting hydrogen profile in such a direction would show a cut-off at the frequency corresponding to the line-of-sight velocity at T.

This principle was used by Kwee, Muller and Westerhout (1954) who determined the cut-off frequencies for a series of galactic longitudes after making allowance for peculiar motions of the gas clouds, which they assumed to be random and with similar dispersions in different parts of the galaxy. These cut-off frequencies were converted to rotational velocities using optical values for R_s (= 8·2 kpc) and ω_s. The latter are shown as points in Fig. 11, taken from Rougoor and Oort (1960), together with corresponding Sydney values for the southern side. The derived rotational curve is drawn as the upper envelope of the observed points because, if a tangential point should fall in a region devoid of gas, e.g. between two spiral arms, the derived velocity would refer to regions further out and would be too low. Those parts of the curve within 3 kpc are derived from evidence discussed in § 5.3, those beyond the sun from optical evidence.

A very fundamental application of the orbital velocity function is to the derivation of the distribution of mass in the galaxy. This is similar in principle to the determination of the masses of the sun and earth by observation of the orbital

periods of their respective satellites, the earth and the moon, but is complicated by the three-dimensional aspects of a rotating galaxy. The standard model of the mass distribution in our galaxy put forward by Schmidt (1956) was derived in this manner. However, the rotation curve used by Schmidt was an earlier version differing from Fig. 11 within 3 kpc of the galactic centre.

It is interesting to note that at the same time 21-cm radio observations were applied by Kerr and de Vaucouleurs (1955, 1956) to determine the first rotational velocity models for the two Clouds of Magellan and to a subsequent determination of the masses of these galaxies.

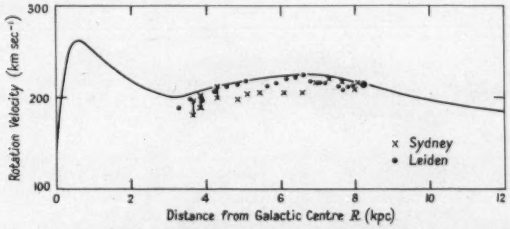


Fig. 11. The variation of rotational velocity in the equatorial plane of our galaxy (according to Rougoor and Oort 1960).

5.2. *Spiral and Other Structure in the Galactic Disk*

The most detailed studies of radio astronomy are those of the distribution of interstellar matter in the galactic disk. Their significance in relation to spiral structure has been most widely discussed, but the problems of the shape of the disk and of the distribution of neutral and ionized hydrogen in it are also of considerable interest.

5.2.1. *Spiral structure.*

The problem of the physical cause of spiral structure observed in so many galaxies is one of the most challenging in astronomy. In general terms it is known that spiral galaxies are highly flattened, fast rotating systems in which interstellar matter (gas and dust) is highly concentrated to the equatorial plane. This matter, however, is not uniformly distributed but shows characteristic but very diverse spiral structure. In association with the gas and dust numerous blue supergiant stars are found, which, because of their enormous energy output, must have a very limited life ($\simeq 10^6$ years) and so must be newly formed. These stars, and the emission nebulae (HII regions) excited by their ultraviolet light, are the prominent sources of light in the spiral arms, but since they are short lived, the gas, and possibly the dust, must be a more basic feature of the spiral structure. One of the most difficult aspects of the problem is the relation to galactic rotation. The period of rotation of objects at different distances from the galactic centre varies grossly. According to the data of Fig. 11 it is 230 million years at the sun, 100 million at 3 kpc from the centre, and 12 million at 500 parsecs. Hence in

periods comparable with the age of the universe, 10^9 or 10^{10} years, spirals would be expected to wind up into tight balls and the arms lose their identity. But typical external spirals show only two or three complete turns. There must therefore be some factor, at present unknown, controlling the arms despite the mass motion of stars and gas which presumably must actually pass through them.

With the evidence of the external galaxies before them it was natural for astronomers to search for signs of spiral structure in our own system, but up till 1951 only indirect evidence had been obtained. In 1951 a breakthrough was made on

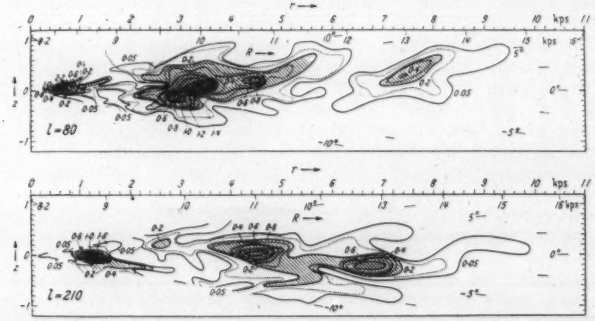


Fig. 12. Sample cross sections of the neutral hydrogen distribution in planes perpendicular to the galactic equator at $l^{\mathrm{I}} = 80°$ ($l^{\mathrm{II}} = 112°{\cdot}3$) and $210°$ ($l^{\mathrm{II}} = 242°{\cdot}3$). Contour numbers are hydrogen atoms per cm³ (Oort 1959).

two fronts. On the optical side Morgan made a systematic survey of O and B supergiant stars and found that they tended to fall into elongated groupings suggesting spiral arms. On the radio side, immediately after the discovery of the 21-cm line by Ewen and Purcell (1951), Muller and Oort (1951) and Christiansen and Hindman (1952) showed that observations of the line could also be used to trace spiral structure.

The radio method, based on distance determinations derived from a galactic rotation model as explained above, was developed by the Leiden group and the results of a very extensive survey of the part of the Milky Way visible from Leiden are described in a series of papers of which those of Muller and Westerhout (1957), Westerhout (1957) and Schmidt (1957 a) are the culmination. Sample cross sections perpendicular to the galactic plane are shown in Fig. 12. A similar survey of the Milky Way visible from Sydney has now been completed (Kerr, Hindman and Gum 1959) using similar reduction procedures. However, the results using the same velocity model gave a distribution which Kerr (private communication) considers implausible. The currently available map of spiral structure based on joint Leiden–Sydney work is shown in Fig. 13. We believe this map to be correct in general features, including the existence of well-defined spiral structure

and the approximate placing of the major arms; we do not know what confidence to place on details. The basic uncertainty lies in the reliance on a velocity model for distance determination, and consequent vulnerability to systematic motions other than simple rotation. Definite results await further study.

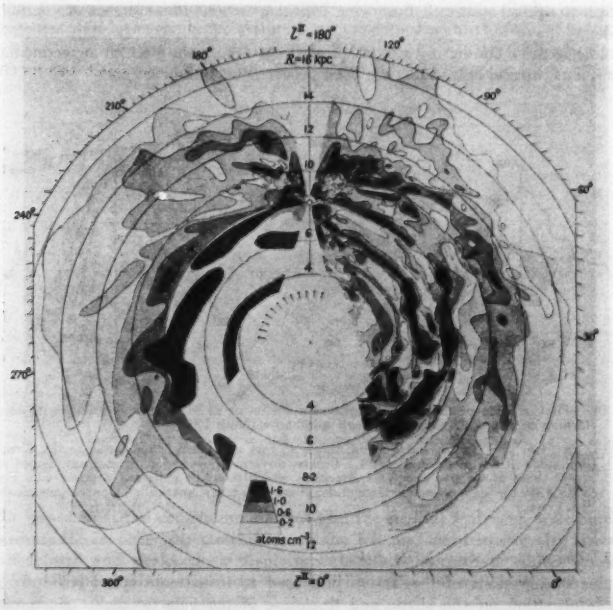


Fig. 13. A provisional Leiden–Sydney map of the distribution of neutral hydrogen in the plane of our galaxy, showing spiral structure (Oort, Kerr and Westerhout 1958). (Coordinates: longitudes l^{I} are shown systematically; for the galactic centre and anticentre directions l^{II} are given.)

New information has recently become available from continuum observations. Mills (1959 b), from his survey at 85 Mc/s with the 'Mills Cross' which has a beam width of 0°·8, has shown that when the effects of discrete sources are subtracted, the variation of intensity around the galactic equator shows a series of steps (see Fig. 14). These, he suggests, are due to the line-of-sight becoming tangential to successive spiral arms. A most interesting point is that a two-start equiangular spiral can be drawn—see Fig. 15—which is tangential to the direction of each one of the steps. The directions along the local arm coincide with two well known maxima, at $l^{I} = 48°$ and 232° in Cygnus and Vela-Puppis respectively. These

are prominent in Fig. 1. Mills suggests that this spiral, which is presumably much more regular than the actual one, may nevertheless be the spiral of best fit.

Because the 21-cm picture is being reconsidered, we shall not make a critical comparison between 21-cm and continuum results. Difficulties also exist with respect to optical studies. Following Morgan, various observations of ' spiral arm markers ' have led to an embarrassing richness of objects through which it is possible to draw conflicting sets of spiral arms. A recent attempt at reconciliation of the data, optical and radio, has been reported by Bok (1959).

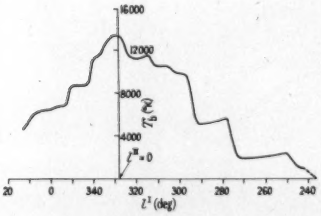


Fig. 14.

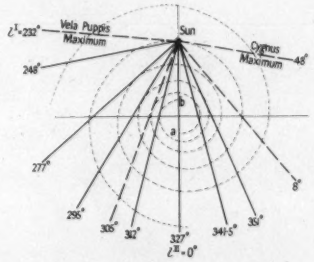


Fig. 15.

Fig. 14. The variation of the disk background radiation at 85 Mc/s around the galactic equator, showing steps probably associated with spiral arms (Mills 1959 b).

Fig. 15. A ' two start ' equiangular spiral drawn so that each turn is tangential to the directions (indicated) of the steps in Fig. 14. Broken direction lines refer to less pronounced steps, except those in Cygnus and Vela directions which point to extended maxima (see Fig. 1) (Mills 1959 b). (Coordinates : longitudes l^{I} are shown systematically ; for the galactic centre l^{II} is given.)

5.2.2. *The shape of the galactic disk.*

The observations of the 21-cm line used to investigate spiral structure also give information on the shape and thickness of the galactic disk. Fortunately, however, the results are much less subject to uncertainties in interpretation. It turns out that the sun is very close to the central plane of the hydrogen layer ; consequently errors in distance estimation, for example due to non-circular velocities, have relatively minor effects on derived displacements of parts of the layer from the mean plane.

The shape of the layer has been carefully studied in connection with the definition of the new system of galactic coordinates by Gum, Kerr and Westerhout (1960) and Gum and Pawsey (1960). The layer is found to be fairly uniformly thick, about 250 pc to half density, from distances of about 3 kpc from the galactic centre to beyond the sun. The inner region is about half as thick (see § 5.3). The displacements of the centre of the layer from the plane of the new galactic equator, which was chosen as the plane of best fit, are shown in Fig. 16. In the region within about 8 kpc of the galactic centre the displacements are surprisingly small ;

over an area 8000 pc in radius very few points depart from the plane by more than 25 pc, a departure of 1 in 300. Outside this region Fig. 16 shows that there are relatively large systematic displacements—the right-hand side is folded upwards, the left-hand side downwards, through distances of 600 or 800 pc.

Only a very small part of the region shown in Fig. 16 is unobscured optically so that the evidence is almost completely from the radio side. Radio continuum observations do however support the 21-cm picture, despite the lack of a distance marker. It is found that, in the general direction of the centre, the continuum observations display a narrow and well-defined ridge of intensity in the direction

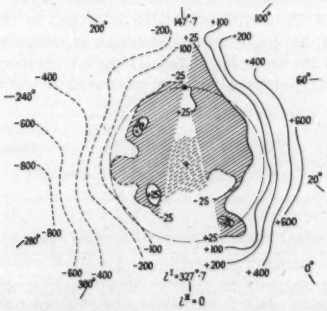


Fig. 16. A ' relief map ' of the galactic disk showing displacements, in parsecs, of the centre of the neutral hydrogen layer from the plane of the new equator (Gum, Kerr and Westerhout 1960). (Coordinates : as in Fig. 15.)

to be expected if the continuum sources and the neutral hydrogen were similarly distributed. In other directions, e.g. to the left and right in Fig. 16, the ridge is ill-defined and peaks show displacements in the directions which would be expected if the continuum sources were distributed in the same way as the neutral hydrogen in the distorted regions.

The high degree of flatness of the inner region must have important implications concerning the dynamics and evolution of the galaxy. The oscillating motion under gravitational forces of gas clouds normal to the plane, i.e. in the z-direction, will be heavily damped by collisions between clouds, so that it is probable that the layer of gas acts as a marker for the surface of zero z-component of gravity. The layer is presumably also normal to the axis of rotation of the galaxy. Since the gas is flat the mass distribution, which is now dominated by the stars, must also show a corresponding regularity.

Irrespective of the evolutionary aspects, however, it is clear that the plane defined by this neutral hydrogen, ' the HI Principal Plane ', must be a fundamental reference plane for galactic studies. For such reasons this plane has been used to fix the equatorial plane of the new galactic coordinate system—the first major excursion of radio observations into the field of positional astronomy.

The departure from a plane in the outer regions which was first pointed out by Kerr, Hindman and Carpenter (1957), and Burke (1957), occurs where the gravitational forces of our own galaxy are very low and suggests the action of external forces. Kerr *et al.* suggested tidal effects due to the Clouds of Magellan but thought them inadequate. An alternative idea was recently put forward by Kahn and Woltjer (1959); they suggested a hydrodynamic effect due to the motion of the galaxy through intergalactic gas—an intergalactic wind effect. There is at present no firm evidence as to the responsible mechanism.

5.2.3. *The distribution of interstellar hydrogen in the galaxy.*

A by-product of the analysis of 21-cm line observations described in § 5.2.1 is the mean distribution of neutral interstellar hydrogen in the galactic disk. The radial distribution of density of neutral hydrogen in the galactic plane as given by Westerhout (1958) is shown in Fig. 17(*a*). Figure 17(*b*) shows similar information derived by Kerr and Hindman (1957) but expressed, not in atoms per cm^3, but as

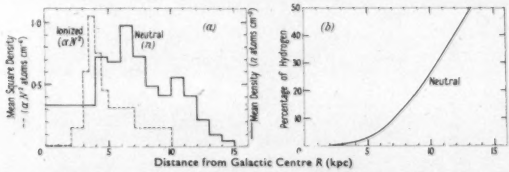


Fig. 17. (*a*) The radial density distributions of neutral hydrogen (mean values) and ionized hydrogen (mean square values) in the galactic plane (Westerhout 1958). (*b*) Neutral hydrogen density expressed as a percentage of the total, stars and gas (Kerr and Hindman 1957).

a fraction of the total density, stars plus gas, in each neighbourhood. In both cases values within 3 kpc of the galactic centre are unreliable in the light of later information (§ 5.3).

Little is known regarding the presence of molecular hydrogen in interstellar space, but ionized hydrogen can be studied using the thermal component of the continuum. Westerhout (1958) estimated the magnitude of the latter at 1390 Mc/s by subtracting estimated values of the non-thermal component derived from 85 Mc/s measurements by Hill, Slee and Mills (1958), adjusted according to an assumed spectrum, from his measured values. From the variation of the equatorial thermal component with longitude, and assuming radial symmetry, he derived a model of the radial variation of thermal emissivity. The emissivity is, assuming constant temperature (see § 3.1), a measure of the mean square density. Consequently deductions as to mean density involve assumptions as the degree of homogeneity. Westerhout supposed that the ionized hydrogen is localized in clouds of uniform density, N atoms cm^{-3}, which occupy a fraction α of the space in the region concerned. Under these conditions the emissivity is a measure of $\alpha N^2 (= N\bar{n}$ where $\bar{n}$ is the mean density) and this quantity is also plotted in Fig. 17(*a*). Westerhout suggests 5 or 10 as a typical value of N. Westerhout's estimates of the masses of neutral and ionized hydrogen (assuming $N = 7$ cm^{-3}) in the whole galaxy are given in Table 1.

The figures in this table refer only to the galactic disk. There is presumably some hydrogen distributed throughout the halo but we cannot yet assess how much.† It appears that the bulk of the hydrogen is neutral and nearly the whole of the matter in our galaxy is now condensed into stars. However, in the outer parts (see Fig. 17(*b*)), there are regions where the masses of stars and interstellar gas are approximately equal. The corresponding slow decay of hydrogen density relative to star density with increasing distance from the centre was first observed in an external galaxy in the two Clouds of Magellan by Kerr, Hindman and Robinson (1954). It appears as a logical consequence of evolution if star formation is strongly favoured by high gas density.

Table 1. Masses of Neutral and Ionized Hydrogen in the Galactic Disk

	Mass	Percentage of whole galaxy§
Neutral	$1{\cdot}4 \times 10^9\ M_\odot$	2
Ionized	$6 \times 10^7\ M_\odot$	0·1

§ Total mass $= 7 \times 10^{10}\ M_\odot$ (Schmidt 1956).

5.3. *Expansion and Other Phenomena near the Galactic Centre*

In the radio continuum there is a strong source, Sagittarius A, in the direction of the galactic centre. The probable identity of this source with the galactic nucleus was pointed out by McGee and Bolton (1954). After some controversy the identification is now generally accepted and the position of the source is now considered to give the most precise estimate of the direction to the galactic centre (Oort and Rougoor 1960).

The source is complex and only stands out strongly above the galactic ridge when observed at decimetre or centimetre wavelengths. This indicates the presence of a thermal source but according to the interpretation by Westerhout (1958) of his own 1390 Mc/s results and 85 Mc/s results by Mills (1956),‡ the whole source appears to consist of a non-thermal source of dimensions $2° \times 1°$, or 300×150 pc, with a peak brightness temperature of 25°K at 1390 Mc/s, together with a thermal source (HII region) of dimensions $0°{\cdot}5 \times 0°{\cdot}25$, 80×35 pc, and brightness temperature 500°K located at the centre. He estimates the emission measure as 360 000 cm^{-6} pc, the mass as $2{\cdot}5 \times 10^5\ M_\odot$ and the density as 85 atoms cm^{-3}. This is easily the most spectacular source in his list.

Drake (1959 b) has made very high resolution observations at 3 cm and finds the central thermal source broken up into two peaks separated by 30 pc. The non-thermal source shows peaks at two points, each 90 pc from the centre, and he suggests that it may actually be in the form of a ring. Thus a definite picture of the radio continuum emission near the galactic nucleus has not yet emerged; future investigations should lead to a clearer picture.

† The chief relevant evidence is that the neutral hydrogen in the whole of the Andromeda nebula is about the same fraction of the whole mass as is the disk hydrogen in our galaxy.

‡ See Hill, Slee and Mills (1958) for corrections to temperature scale.

Turning now to 21-cm studies, recent observations have led to the discovery of quite unexpected large-scale gas motions in the vicinity of the galactic nucleus—a discovery which is leading to an optical search for similar phenomena in external galaxies. This is in contrast to the work on the spiral structure over the main part of the disk, where radio observations give a comprehensive picture of a structure which was expected to exist from observations of external galaxies.

The first evidence for a peculiarity in the central region was obtained by Kwee, Muller and Westerhout (1954) when they observed faint wings on profiles taken in the galactic plane within about 20 degrees of the galactic centre (see Fig. 10). These wings showed velocities far in excess of anticipated rotational velocities and they were tentatively attributed to turbulence in this region. Later observations by van Woerden, Rougoor and Oort (1957) showed that the wings were not irregular, as they would be if due to turbulence, but included peaks which showed continuity over a wide range of longitudes. One such series of peaks could be followed over more than 10 degrees. At the distance of the galactic centre, 10 degrees corresponds to an extent of the source of about 2 kiloparsecs. It was concluded that the peak was due to an extended mass of gas like a spiral arm and, for the reason given below, the mass was designated the '3 kpc arm'. This arm passes across the longitude of Sagittarius A and, where it crosses, the radiation from the latter is strongly absorbed. The arm is therefore between us and the galactic centre. The astonishing feature of the arm is that, in the direction of the centre, the line-of-sight velocity is not near zero, as it would be for simple rotation, but is -50 km/sec. At this point there must be an expansion component of this order. There is no published evidence for such phenomena for arms further out.

More extensive observations have now been completed (Rougoor and Oort 1960), and show a composite picture of expansion and rotation in the central region. The gas in the central region appears to be concentrated in a layer about 120 pc in thickness (to half density), about half the thickness of the layer in the outer region. The important features of the observations are shown in Fig. 18, which gives the line profiles along the actual galactic equator ($b^{II} = 0°$) in the form of contours of brightness temperature as a function of Doppler velocity V, relative to the local standard of rest, and galactic longitude (l^{II}). The wide, nearly vertical, strip of intense radiation passing through the centre is due to hydrogen along the whole length of the sun-centre line and has been omitted. There is a narrow valley, at negative velocities, in the direction of Sagittarius A ($l^{II} = 0°$) which is only partly shown. This valley is only apparent, being due to absorption of Sagittarius A emission. All frequencies corresponding to negative velocities show it; positive ones do not. This is clear evidence for outward flow of gas from the centre.

Before interpreting Fig. 18 let us consider the form of the (l^{II}, V) diagram to be expected from a ring of gas concentric with the galactic centre. If it were in simple rotation it would give a ridge of high intensity in the form of a straight line through the origin, the quadrants depending on the sense of rotation; if in pure expansion, an ellipse with axes on the l^{II} and V axes. Combined rotation and expansion would yield a tilted ellipse. Now returning to Fig. 18, we see two such ridges, ABC and EFGH. ABC is that associated with the '3 kpc arm'. It appears to be a portion of a tilted ellipse which seems to be traceable on Sydney observations down to about $l^{II} \simeq -25°$, in which direction it is 3 kpc from the centre.

Much of the remainder of the ellipse, even if it were complete, would be obscured by the wide strip of intense radiation. The appearance of the ridge thus suggests the existence of a part of a ring of hydrogen, or a spiral arm, showing both expansion and rotation. Values of expansion and rotation velocity derived from the intersection and slope of the ridge at $l^{II} = 0°$ are 50 and 200 km/sec respectively.

The ridge EFGH is a reasonable approximation to a straight line through the origin and is interpreted as due to a ring of gas of radius 600 pc and having a pure rotational velocity of 200 km/sec. Among other broad features of the distribution, the tendency to concentration in the first and third quadrants suggests widely

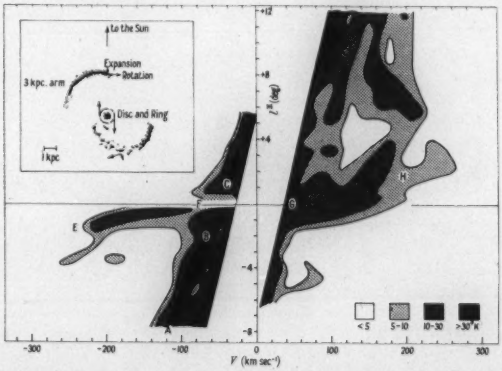


Fig. 18. Contours of H-line brightness temperatures along the equator near the galactic centre as a function of galactic longitude and Doppler velocity. The inset sketch shows the derived distribution of neutral hydrogen near the centre (Rougoor and Oort 1960).

distributed matter with a combination of rotation and expansion motions, and the marked asymmetry of the distribution must indicate a corresponding asymmetry in the velocity–density distribution.

As a result of more detailed analysis Rougoor and Oort found evidence for a small, very dense disk of neutral hydrogen at the centre. These features are indicated in the sketch in the upper left corner of Fig. 18 which is copied from one due to Rougoor and Oort. They emphasize the sketchiness of the positional evidence excepting that relating to the central disk and ring.

Radio observations have then supplied evidence for several things quite new to astronomy. First comes the expansion observed somewhere in the region between 1 and 4 kpc from the centre. It is not known whence comes the gas nor where it goes. Rougoor and Oort estimate the amount flowing out to be about one solar mass per year. It must either come from gas expelled by stars or flow in from somewhere, perhaps the corona. In either case the mere fact of radial flow must be a major point in understanding the dynamics of galaxies. Next comes the peculiar ring and disk structure at the centre which is completely unexplained.

Finally there is the fact that in the very centre of our galaxy, a region which according to conventional ideas is devoid of gas, we find the exceedingly dense disk of neutral hydrogen and what is apparently the most intense 'HII region' type of source in the whole of our galaxy. Are radio astronomers being tricked by a new type of source whose spectrum simulates that of an HII region? Does our galaxy have HII regions where external galaxies have none? Or have optical astronomers been blinded by the intense light of the stars in the nuclei of other galaxies?

Such are the provocative questions raised by the radio observations and the challenge is already being taken up by optical astronomers. Münch (1960) has reported studies of the motion of interstellar gas, using the OII doublet at $\lambda = 3726$–3729Å, in the central regions of the Andromeda nebula. He has found large velocities of the order of 50 or 100 km sec^{-1} in the radial direction, presumably outwards, at distances of from 200 to 400 pc from the nucleus. These observations conform with the radio ones as regards the existence of radial flow. They differ in that the optical ones refer to ionized gas, the radio ones to neutral, and the region observed is much closer to the centre in the optical case.

5.4. *The Galactic Corona*

Galactic radiation so far discussed in §§ 5.1, 5.2 and 5.3 all originates within a relatively thin layer about the galactic plane. There is, however, another interstellar component of the galaxy, roughly spherical in shape and centred on the galactic nucleus and completely surrounding these disk components. This component was first discovered by Shklovsky (1952) from a careful analysis of radio surveys of the sky at various frequencies. We now call this component the galactic corona, or halo—its radio emission is non-thermal. Shklovsky was unable to ascertain the origin of the radiation until Pikelner (1953), on cosmic ray evidence, postulated chaotic magnetic fields associated with the galaxy existing far from the galactic plane. Shklovsky (1953 c) then proposed that the emission was synchrotron radiation produced by relativistic electrons moving in magnetic fields. It is interesting to note that galactic radio emission had previously been discussed in similar terms by Kiepenheuer (1950), Ginzburg (1951) and Getmansev (1952); they, however, had not emphasized the existence of a corona. This hypothesis immediately linked radio emission with the existence of cosmic rays in the galaxy.

The existence of a corona in our own and other galaxies is now accepted and implies the existence of high energy electrons and magnetic fields. Because the radio spectrum is of the form $b \propto f^{-0.5}$, approximately, the energy spectrum of the electrons (see § 3.2) is of the form $E^{-2.0}$ which is similar to that of cosmic rays. The emission is then proportional to $kH^{1.5}$. Moreover, there are probably constituents present in addition to electrons and magnetic fields, e.g. protons and the heavier nuclei of cosmic rays, and matter related to the magnetic fields. What these additional components are, and the manner in which they fit into the large scale processes occurring in the galaxy, are major problems concerning the corona. We now outline the principal observational data available.

The survey of radio emission over the sky made by Mills (1959 b) at 85 Mc/s is very well suited to the study of non-thermal emission from the galaxy. For at this frequency the distribution of non-thermal radiation is least affected by emission

PLATE I

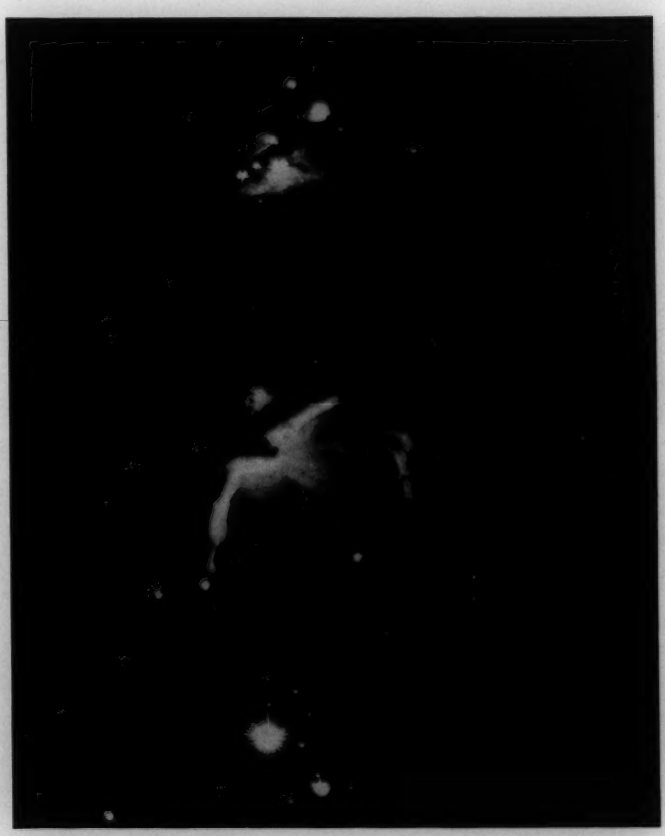


Fig. 7. The Orion nebula—an emission nebula (a Lick Observatory photograph).

PLATE II

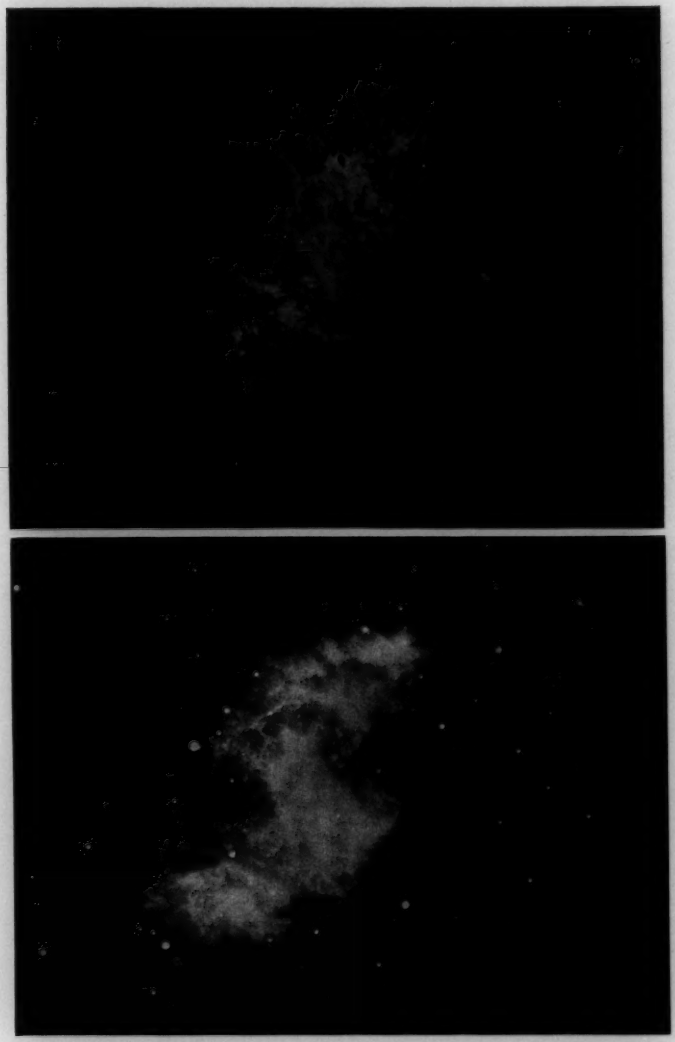


Fig. 8. The Crab nebula—a supernova remnant (radio source Taurus A). The upper photograph shows the filamentary shell, the lower shows the amorphous inner mass. (*By courtesy* J. H. Oort.)

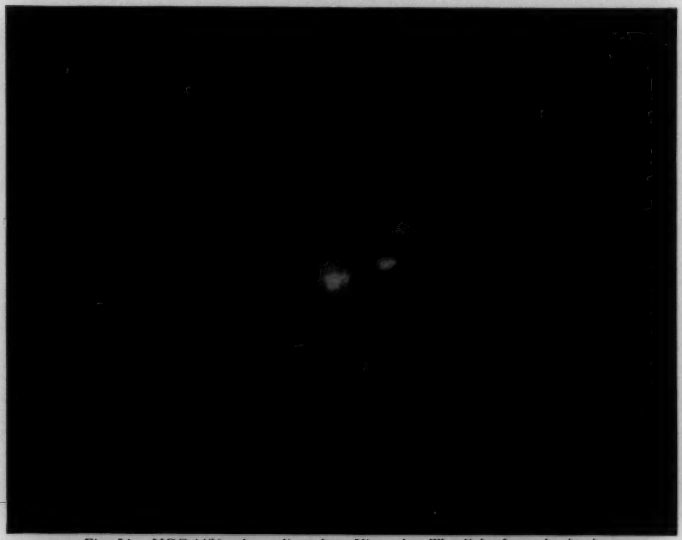


Fig. 24. NGC 4486—the radio galaxy Virgo A. The light from the jet is partially plane polarized. (*By courtesy* R. Minkowski.)

Fig. 27. NGC 5128—the radio galaxy Centaurus A. (*By courtesy* R. Minkowski.)

PLATE IV

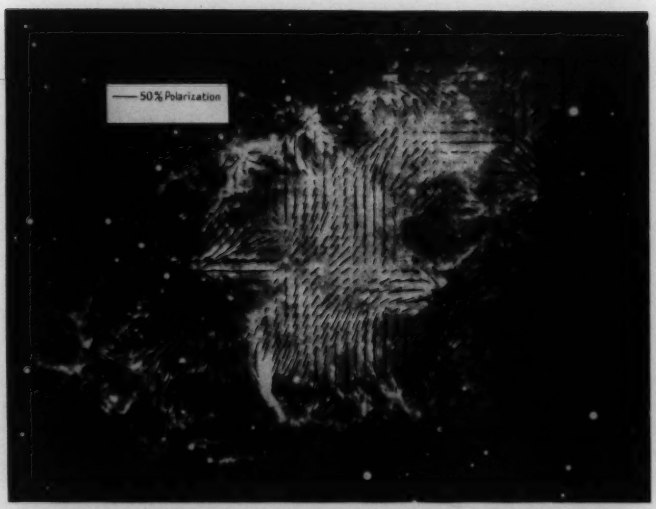


Fig. 9. The degree and direction of polarization of light from the amorphous part of the Crab nebula, superimposed on a picture showing the filaments (Woltjer 1958).

from, or absorption by, HII regions and equally significant is the very high resolution of the survey. Several scans taken from this survey across the galactic plane (Fig. 19) show the galactic corona clearly as a wide distribution in which brightness falls slowly away from the galactic plane. Superimposed is the very narrow band of the non-thermal disk component. The general features characterizing the corona as seen on the sky are shown by the contours from this survey in Fig. 20; this illustrates clearly the association of the corona with the galaxy. It must be remembered, however, that there is presumably an isotropic extragalactic component included in these temperatures amounting to perhaps 450°K.

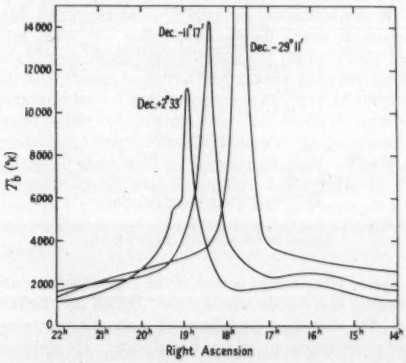


Fig. 19. 85 Mc/s Mills Cross scans (epoch about 1956) across the galactic equator at different declinations. Note the narrow ridge due to the disk superimposed on the broad distribution due to the corona.

Mills has analysed this material with a view to obtaining the spatial distribution of emissivity (emission per unit volume) of the coronal radiation on the assumption of radial symmetry about the galaxy's axis of rotation. The radial distribution of emissivity in the galactic plane so obtained is shown in Fig. 21. Within the sun's distance from the galactic centre the emissivity distribution is defined uniquely by the longitude distribution of brightness and this is shown by the full curve. Outside this domain, however, a unique solution is not possible. The broken curve therefore represents only one of the possible distributions compatible with the observations. After subtracting the estimated extragalactic component, the latitude intensity distribution may be fitted very satisfactorily by a set of homogeneous oblate spheroids derived from the radial distribution and having an axial ratio of 1·45. Thus the coronal emission shows a concentration to the galactic centre and to the galactic plane. Further, Mills finds the emissivity to be 5×10^{9} watt pc^{-3} $(c/s)^{-1}$ at the centre and the total coronal emission $\simeq 10^{22}$ watt $(c/s)^{-1}$, which is about 10 times the non-thermal radiation issuing from the disk component. Broadly similar results were obtained by Baldwin (1955) on the basis of the best

fitting of several arbitrary models. Nearby spiral galaxies show similar coronas (see § 6.1).

Optical evidence indicating the existence of gas clouds at distances from the galactic plane of 500 or 1000 pc has been invoked by Spitzer (1956) as further

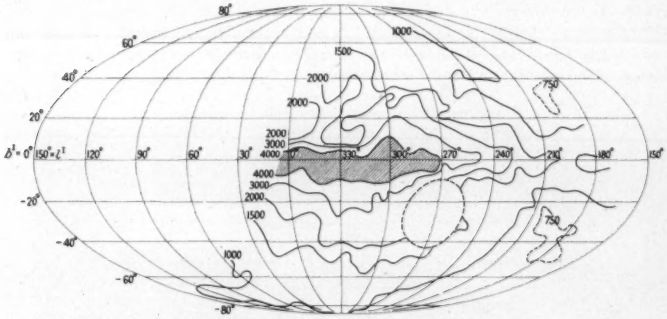


Fig. 20. The brightness distribution of the galactic coronal emission derived from the Sydney 85 Mc/s survey (Mills 1959 b).

evidence for a corona. His concept is that, if the regions surrounding these clouds were effectively empty, the clouds would have ' boiled off ' before reaching their present positions. He therefore postulates an exceedingly tenuous atmosphere, but with a pressure adequate to retain the clouds. He suggests a completely

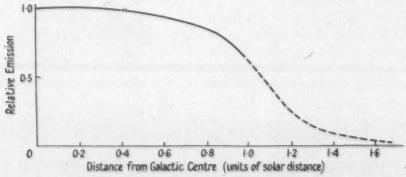


Fig. 21. The radial distribution of coronal emissivity in the galactic plane calculated assuming radial symmetry (Mills 1959 b).

ionized gas of density 5×10^{-4} atoms cm^{-3} and $T = 10^6$ °K which is supported by its own pressure and extends over the region of the observed corona. High energy electrons and magnetic fields do not appear explicitly in this hypothesis of a static corona which is a very considerable extrapolation from phenomena observed on the fringe of the galactic disk.

The dynamical model proposed by Pikelner and Shklovsky (1957, 1958, 1959) stems from the synchrotron interpretation of the radio emission. The corona is supposed to consist of irregular clouds of gas in violent motion (100 km sec^{-1}) in a

medium containing magnetic fields and high energy electrons and atomic nuclei (cosmic rays). The gas density is approximately 3×10^{-3} cm^{-3} (relativistic electrons $< 1\%$ of this) and the magnetic field approximately 3×10^{-6} oersted. Macroscopic motions are supposed to play a part in maintaining the magnetic fields and in accelerating high energy particles, through Fermi-type processes. Supernovae are considered to be the primary source of high energy particles.

These hypotheses are, in view of the extreme paucity of data, very highly speculative and observational tests would be most valuable. Polarization of the radio emission, if established, would provide strong confirmation of the synchrotron theory. Razin (1958) has attempted such observations and claims to have observed, at a wavelength of 1·45 m, about 3% linear polarization over all parts of the sky observed excepting near the galactic equator. These measurements are very difficult and confirmation of this most important result would be very desirable. Thompson (1957) and Pawsey and Harting (1960) have attempted to check the results but without success. That the polarization present should be small is not surprising, as Razin points out, because there are likely to be strong depolarizing influences present arising from the chaotic nature of the magnetic field in emitting regions along any line of sight, and to Faraday rotation. The relativistic electron content of the primary cosmic rays is another basic factor in the theory but is difficult to measure and positive results have not yet been obtained. A further possible observation is the detection of 21-cm line emission from the corona of our own or external galaxies.

§ 6. Comparative Radio Observations of Normal Galaxies

Observations of external galaxies yield information of two sorts—they may give a better insight into phenomena which are observed in our own galaxy, and they may show the way in which particular features vary from one galaxy to another. In the first category the extent of the coronas of external galaxies can be observed directly, whereas from within we can only infer the extent of the galactic corona from indirect evidence. Indeed, in his historic 1952 paper on the corona of our galaxy, Shklovsky cited as vital evidence the great angular extent of the radio emission from the Andromeda nebula, as observed by Brown and Hazard. In the second category, one may cite the observed dependence of the interstellar hydrogen content on galactic type—this is one of the most important facts relating to the evolution of galaxies.

Because of small angular sizes this is a field in which radio limitations are very severe, and it has so far been possible to observe radiation from only a few dozen external galaxies, and to distinguish any detail at all in only half a dozen. We shall first discuss continuum evidence on coronas of galaxies, and then 21-cm line evidence on hydrogen content of galaxies.

6.1. *Continuum Observations*

The radio emission from a few dozen of the optically bright galaxies has been examined, principally by Brown and Hazard using the Manchester 220-foot fixed parabolic reflector, and by Mills and his colleagues using the Sydney 85 Mc/s Mills Cross. The results of these studies of normal galaxies are given in Table 2

taken mainly from the compilation by Mills (1959 d). This list excludes galaxies with abnormally high radio emission.

It appears that all 17 of the bright galaxies of a late type (Sb and later) which have been investigated are detectable radio emitters; none of 'earlier types' has been detected. Quantitatively the three irregular galaxies all have a ratio of radio to

Table 2. Continuum Observations of Normal Galaxies

Galaxy	Class	$m_{1\cdot9}$	m_p	$m_{1\cdot9}-m_p$
LMC	Irr	3·4	0·6	2·8
SMC	Irr	5·5:	2·5	3·0:
M 31†	Sb	5·8	4·0	1·8
M 33†	Sc	7·8	6·2	1·6
NGC 4945	SBc	8·7	7·8:	0·9:
NGC 1068	Sb_p	8·7	9·6	−0·9
NGC 5236	SBc	8·8	7·5:	1·3:
NGC 253	Sc	9·0	7·5	1·5
NGC 3031	Sb	8·9	7·7	1·1
NGC 891	Sb	9·0:	10·7	−1·7:
NGC 5194/5	Sc	9·7	8·5	1·2
NGC 4258	Sb	9·8	9·1	0·7
NGC 300	Scd	9·9:	8·5:	1·4:
NGC 6744	SBbc	10·2	9·1	1·1
NGC 2841	Sb	10·4:	10·2	0·2:
NGC 55	Irr	10·9	7·8	3·1
NGC 5457	Sc	> 9·1	8·1	>1·0
NGC 205	E5	> 9·0	8·9	>0·1:
NGC 221	E2	> 9·7	9·1	>0·6
Fornax	dE	>10·0	7·5:	>2·5:
IC 5267	Sa	>10·9	10·8	>0·1
Sculptor	dE	>11·1	7·1:	>4·0:
NGC 1291	SBO	>11·2	9·4	>1·8
NGC 4594	Sab	>11·3	8·9	>2·4
NGC 3115	E7	>11·4	10·15	>1·2

† Some data for this galaxy taken from Brown and Hazard (1959).

: Less reliable measurements.

light emission given by $m_{1\cdot9}-m_p \simeq 3$;‡ a considerable group of Sb and Sc galaxies have values close to +1. The earlier types appear to have a smaller ratio of radio to light emission but the sample is marginal. It is a surprising result, not yet explained, why there should be such a small spread in the ratio of light to radio emission in the I and the Sc and Sb classes.

For four of the normal galaxies, the two Clouds of Magellan, M 31 and M 33, maps of the radio brightness distribution have been prepared. Of these, M 31 is of the most interest because of its similarity in type (Sb) to our galaxy. The observation of this galaxy in 1950 by Brown and Hazard (1951) was the first definite

‡ For easy comparison of photographic magnitudes m_p with radio flux densities S_λ at wavelength λ, a radio magnitude m_λ has been introduced, and is defined by

$$m_\lambda = -53{\cdot}4 - 2{\cdot}5 \log_{10} S_\lambda.$$

The arbitrary constant, −53·4, was chosen to make $m_p \simeq m_\lambda$ at $\lambda = 1{\cdot}9$ m for a normal Sb galaxy. $m_\lambda - m_p$ is then a logarithmic measure of the ratio of radio to light flux received.

radio observation of a normal external stellar system, and their work provided the first brightness contours over a discrete source. Fig. 22(*a*) shows higher resolution contours obtained recently by Large, Mathewson and Haslam (1959) at 408 Mc/s with the 250-foot Jodrell Bank paraboloid over the region of M 31. The irregularity of the general field, presumed due to foreground irregularities in our galaxy, is obvious from the irregularities in the outer parts of the figure. Fig. 22(*b*) shows an earlier, and lower resolution, contour map prepared from 158 Mc/s observations on the same aerial by Brown and Hazard (1959), in which the estimated effects of the foreground have been removed. The outstanding feature of these

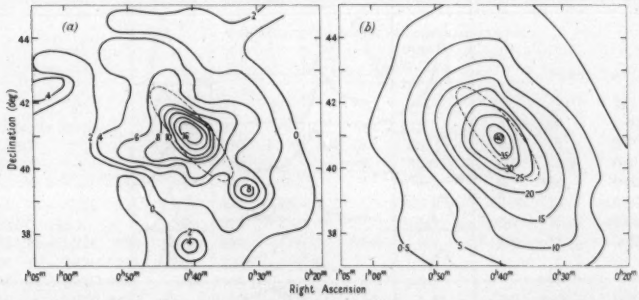


Fig. 22. Radio contours of the Andromeda nebula (M31): (*a*) as observed at 408 Mc/s with a beam width of about 0°·8 (Large, Mathewson and Haslam 1959), and (*b*) at 158 Mc/s after removal of the estimated foreground—the beam width is about 1°·9. The broken ellipse encloses 95% of the optical luminosity (Brown and Hazard 1959).

observations is the very great extent of the radio source in relation to the optical nebula. The contour enclosing 95% of the optical luminosity is shown dotted in Fig. 22(*a*) and (*b*).

More detailed study of the observations at both frequencies showed evidence of two superimposed distributions corresponding to the disk and corona of our galaxy. At 158 Mc/s the disk of M 31 contributes about 10% of the radio emission, the corona the remaining 90%, agreeing closely with Mills' (1959 b) estimates for our galaxy, see § 5.4. Similar studies were made of the Type Sc galaxy, M 33, and some of the features of the two galaxies as observed by Brown and Hazard (1959) are summarized in Table 3.

Table 3. Radio Sizes and Shapes of M 31 and M 33

Galaxy	Size†		Ratio of diameters (radio/optical)	Linear extent of radio source (kpc)
	Radio (deg)	Optical (deg)		
M 31	10·6 × 6·1	3·6 × 1	3	120
M 33	5·4 × 2·4	1·4 × 1	4	60

† Total widths between 1% contours.

Similar observations have been made for the two Clouds of Magellan by Mills (1955). In these irregular galaxies Mills did not find evidence for disk and corona sub-systems, but he did find the familiar extension well beyond the optical limits. It now seems fairly certain that a vast corona, which we assume to be evidence for energetic electrons and magnetic fields, is a normal feature of late-type galaxies.

6.2 *Hydrogen Line Observations*

The first 21-cm observations of external galaxies were those of the Clouds of Magellan by Kerr, Hindman and Robinson (1954). The main feature shown by

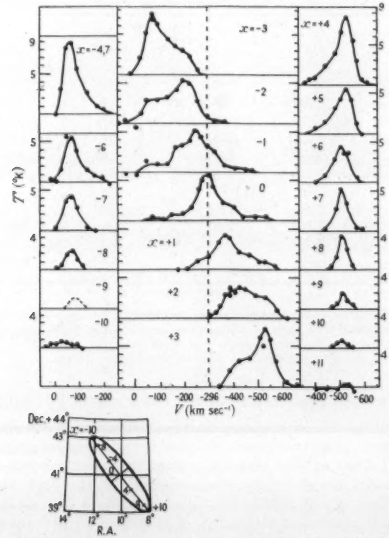


Fig. 23. H-line profiles observed by van de Hulst, Raimond, and van Woerden (1957) on the major axis of the Andromeda nebula at various distances x ($\frac{1}{4}°$ units) from the centre (see insert). (From the book by Steinberg and Lequeux (1960).)

the observations is that the stellar systems are embedded in vast clouds of neutral hydrogen.

Since 1954, observations of 14 other galaxies have been made, mainly by the Leiden group on the 25-metre Dwingeloo paraboloid. Fig. 23 shows a series of profiles obtained at points spaced at $\frac{1}{4}°$ intervals along the major axis of the Andromeda nebula. The contours are centred on an approach velocity of -296 km sec^{-1} with respect to the local standard of rest, and they show very clearly the effect of rotation in the displacement of the whole profile to positive and

Vol. 24
1961

negative velocities for points on the respective sides of the origin. From such studies the authors derived the rotation law and the distribution and total mass of interstellar hydrogen in the nebula. From the former Schmidt (1957 b) has derived the total mass and its distribution. The results of this and similar observations are summarized in a paper by Volders and van de Hulst (1959) from which Table 4 is taken.

It should be mentioned that there is some doubt about the accuracy of the masses in Table 4, particularly for the Clouds of Magellan, but tentatively accepting the figures, the table shows a most interesting trend. The ratio of hydrogen declines

Table 4. 21-cm Observations of Normal Galaxies

	General data			Radio measurements		Mass		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Galactic System	Sb	—	—	—	125	1·5	100	0·015
LMC	Irr	50	+295	15	30	0·7	1	0·7
SMC	Irr	50	+172	9	30	0·5	1·3	0·4
M 31 = NGC 224	Sb	630	−296	5	8·1	4·5	480	0·009
M 32 = NGC 221	E2	630	−200		⩽0·2	⩽0·06	14	⩽0·004
NGC 205	SBo	630	−236		<0·5	<0·1	18	<0·005
M 33 = NGC 598	Sc	630	−176	1·5	6·3	1·0	18	0·05
M 51 = NGC 5194	Sc	1200	+449		0·5			
M 81 = NGC 3031	Sb	2600	− 45	0·3	0·5		120	
M 101 = NGC 5457	Sc	2600	+253	3·3	1·4	3·0	140	0·02
NGC 253	Sc	2300	− 88		0·4	⩾0·07		
NGC 4236	Sc	3300	+ 33		0·4	⩾0·98		
IC 1613	Irr	440	−242	⩾0·5	1·4	>0·02	0·05	>0·4
NGC 6822	Irr	300	− 23		3·0	>0·03		

Key.

(1) System ; (2) type ; (3) adopted distance (kpc) ; (4) velocity relative to the local standard of rest defined by the stars near the sun (km sec^{-1}) ; (5) radio size (deg) ; (6) max. T_{ant} (°K) ; (7) hydrogen mass (10^9 $M_\odot$); (8) total mass (10^9 $M_\odot$); (9) hydrogen/total.

steadily with galactic type, from about 0·5 for irregulars to 0·05 to 0·02 for Sc, 0·015 to 0·009 for Sb, and falling below the observational limit for earlier types. Further, in individual galaxies hydrogen is considerably more widely dispersed than the stars. The two sets of facts lead to a picture of evolution of galaxies in close conformity with current ideas. In E-type galaxies the process of formation of stars from primordial hydrogen is virtually complete, in irregular galaxies it is half complete, and Sb and Sc spirals are at an intermediate stage. The data do not answer the question as to what causes this difference. Is it because, as the steady-state theory suggests, the irregulars are newly formed galaxies—or that in such small galaxies star formation proceeded relatively slowly? These questions cannot yet be answered, but the collection of data such as shown in Table 4 for a more representative sample of galaxies should supply vital evidence on this fundamental question of cosmology.

§ 7. Radio Galaxies and Cosmology

About a year after the discovery of the first discrete source, Cygnus A, two of the half-dozen sources then known were tentatively identified by Bolton, Stanley and Slee (1949) with two somewhat peculiar external galaxies: NGC 4486 and NGC 5128. These identifications have since been confirmed. Some years later Mills (1952) from a study of the distribution of discrete sources in direction and intensity concluded that they fell into two classes, Class I consisting of galactic objects located near the galactic plane, and Class II of objects isotropically distributed, and being either extragalactic, or galactic and relatively close. It is now clear that a large proportion, if not all, Class II objects are external galaxies, of which NGC 5128 and 4486 are typical. Most of them emit far more intensely in the radio continuum than normal galaxies and these are termed ' radio galaxies '; the above two galaxies emit approximately one thousand times as much as our own galaxy.

An exciting new development came with the discovery by Baade and Minkowski (1954) of a most abnormal galaxy precisely in the position of the Cygnus A radio source. This galaxy appeared to be double and the spectrum showed intense emission lines and a spread in velocities of some thousands of kilometres per second. They concluded the object consisted of two spiral galaxies in a face-to-face collision. Red-shift measurements showed that this galaxy is at the enormous distance of 90 Mpc,† an astonishing result for the second strongest radio source in the whole sky. The total emitted radio power is of the same order as the emitted light. From the cosmological point of view the implication concerning the possibility of observing other similar sources at distances far beyond the optical limits opened up new horizons. On an instrument such as the Sydney 85 Mc/s Mills Cross, such a source should be visible out to a distance where the recession velocity is about $0{\cdot}8\,c$. However, it was not clear how to make use of this opportunity. Many of the unidentified faint sources observed might indeed be gigantic emitters at vast distances, but there was nothing to distinguish them from puny ones relatively close.

At this stage radio astronomy made two false starts. On the one hand Lilley and McClain (1956) reported that they had observed the 21-cm line in absorption in the Cygnus source, red-shifted by an amount corresponding to the optical red-shift. McClain (1960) has since reported his inability to reproduce this result, nor have other observers been successful; so the red-shift method of distance measurement which promised to become available to radio astronomers has not yet materialized. On the other hand the Cambridge group drew unwarranted conclusions from a survey of radio sources in which they had catalogued about 2000 sources (Shakeshaft, Ryle, Baldwin, Elsmore and Thompson 1955). Their catalogue showed a statistical excess of faint sources over that to be expected for a uniform spatial distribution (see § 7.2), and they attributed the excess to an increase in average density, or emission, of radio galaxies lying beyond the borders of the optically observed universe. This conclusion flatly contradicts a basic postulate of the steady-state theory: large-scale uniformity in space. However, the reliability of the observations was challenged by the Sydney group and the objections appear now to be accepted by the Cambridge group (Edge, Shakeshaft,

† Distances in this paper, when derived from red shifts, are based on a value for the Hubble constant of 180 (km sec^{-1}) Mpc^{-1}.

McAdam, Baldwin and Archer 1959). Despite these false starts the challenge remains; it is very probable that a considerable number of the faint sources are beyond the current optical horizon. We now examine the current evidence on the nature of radio galaxies and possible cosmological implications.

7.1. *The Nature of Radio Galaxies*

The most direct evidence on the nature of radio galaxies comes from the eight well established optical identifications. These are listed in Table 5 with some pertinent data.

An essential fact evident from a study of this regrettably small sample is that radio galaxies do not form a homogeneous class. The ratio of radio to optical emission, as given by the column $m_{1\cdot9}-m_p$, ranges from -13 to -4 magnitude, a range of nearly 10^4 to 1. If we include normal galaxies this range is even greater and, except for the group of normal Sb and Sc galaxies having $m_{1\cdot9}-m_p \simeq 1$, there appears to be a continuous distribution of ratios with numbers decreasing rapidly as the radio emission increases.

Similarly the appearance of the galaxies in Table 5 is diverse. A high proportion appear to be cases of collisions, but other galaxies showing apparently similar optical features are not strong radio emitters. One galaxy, NGC 4486 (Virgo A), shows a unique feature, the well-known 'jet' with its linear optical polarization (see Fig. 24, Plate III). Others have shown no recognizable optical peculiarities. All are giant galaxies with absolute photographic magnitudes in the vicinity of -20.

A most interesting feature of the observations is the consistent way in which the radio sources extend over much greater areas than do the optical sources. This feature corresponds to the existence of a corona in our own and other normal galaxies. According to column (11), the typical ratio of radio to optical extent is about 3. Further, the radio and optical brightness distributions tend to be very different.

With the completion of recent surveys of radio sources further attempts at optical identifications have been made by Dewhirst (1959), who used the 3C Cambridge survey, and by Mills (1960), who used the Sydney Mills Cross survey supported by angular size observations. Dewhirst listed eight new probable, or possible, identifications with external galaxies. The majority of these are apparently normal E-type galaxies appearing in pairs or small clusters, and having ratios of radio to optical emission given by $m_{1\cdot9}-m_p$ between -10 and -6. The physical significance of this selection is a mystery but the recognition of a distinct feature is important.

Mills listed 46 probable identifications with galaxies out of 400 sources examined. The galaxies were of a wide variety of types but again a considerable number occurred in pairs or small groups. The important step forward made by Mills was to use his results to estimate provisionally the statistical distribution of the ratio of radio to optical emission, given by $m_{1\cdot9}-m_p$ $(=\underline{m})$, for sources chosen at random from the sky (see Fig. 25). The importance of this distribution lies in the fact that it is the initial step in the determination of the radio luminosity function of galaxies, i.e. the distribution of absolute radio luminosities of galaxies in a volume of space. Mills also made a tentative estimate of this function on the assumption

Table 5. Characteristics of

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Cyg A	19N4A	—	90	2·1	15·1	−13·0	0·66	$\frac{1}{2}' \times 2'$ Two distinct peaks	$18'' \times 30''$	4
Her A	16N0A	—	140	5·0	18	−13	1·3	$2\frac{1}{2}'$	$\frac{1}{2}'$	5
Hyd A	09S1A	—	90	5·3	16	−11	1·0	$1\frac{1}{2}'$	$\frac{1}{2}'$	3
—	16N4A	6166	—	7·4	13·9	− 6·5	1·6	—	$\simeq 1'$	—
Per A	03N4A	1275	40	7·0	12·0	− 5	0·9	$2'$	$\frac{1}{2}'$	4
Vir A	12N1A	4486	7	3·9	9·9	− 6	0·74	$5' \times 2\frac{1}{2}'$ With possible 1° extension	$5'$	1 or possibly 12
Cen A	13S4A	5128	1	2·6	6·5	− 3·9	0·66	$7' \times 3'$ double peak with $7° \times 2°$ halo	$\frac{1}{2}°$	$\simeq 10$
For A	03S3A	1316	10	4·9	9·5	− 4·6	—	$\simeq 1°$	$\frac{1}{2}°$	3

Key.

(1) Radio source; (2) IAU number; (3) NGC number; (4) distance (Mpc); (5)–(7) apparent magnitudes; (5) $m_{1\cdot9}$; (6) m_p; (7) $m_{1\cdot9} - m_p$; (8) spectrum index α; (9)–(11) angular extent; (9) radio; (10) optical; (11) radio/optical; (12) source; (13) linear dimensions (kpc); (14) T_b (3·5) (°K); (15) radio luminance P; (16) M_p; (17) characteristics of galaxy.

Notes. (Cols.)

(4) Except for Cen A, see Wade (1959 b), distances are from Ryle (1958), who uses for the Hubble constant 180 km sec^{-1} (Mpc)$^{-1}$.

(5) Radio magnitude from Mills (1959 d). Defined for 1·9 metres by $m_{1\cdot9} = -53\cdot4 - 2\cdot5 \log_{10} S_{1\cdot9}$, where $S_{1\cdot9}$ is the flux density at 1·9 m. Some values are estimated from 3·5 m observations corrected by +0·5 magnitude assuming a spectral index −0·7.

(6) Photographic magnitude from Mills (1959 d) excepting NGC 6166, see Dewhirst (1959). Cygnus A value corrected for absorption.

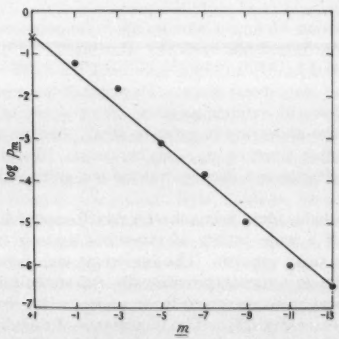


Fig. 25. A relative radio to optical luminosity function for galaxies. $p_{\underline{m}}$ is the probability that a galaxy taken at random shall have a value of $\underline{m}$ between $\underline{m} \pm \frac{1}{2}$, where $\underline{m} = m_{1\cdot9} - m_p$ (Mills 1960).

Eight Radio Galaxies

(12)	(13)	(14)	(15)	(16)	(17)
Cyg A	50×12	10^9	8×10^{36}	$-19{\cdot}9$	Peculiar. Face to face collision—Intense emission lines—In small cluster.
Her A	100	10^7	2×10^{36}	$-18{\cdot}2$	Peculiar. Probable collision.
Hyd A	38	10^7	5×10^{35}	$-18{\cdot}9$	E. Double.
(6166)	—	—	—	$-19{\cdot}5$	E. Multiple—Apparently normal.
Per A	20	10^8	10^{34}	$-19{\cdot}6$	Peculiar. Probable collision.
Vir A	10×5 or possibly 100	10^7 or 10^8	10^{34}	$-19{\cdot}6$	E. With unique 'jet' showing linear optical polarization. Jet would be invisible if much farther away.
Cen A	100×30 whole source	10^4, 10^7 in peak	10^{33}	$-18{\cdot}5$	Peculiar. Possible collision, SO and late type spiral.
For A	$\simeq 200$	10^4	10^{34}	$-20{\cdot}5$	SO. Apparently normal.

Notes. (Cols.)

(8) Defined by $S_f = S_0 f^{-\alpha}$. Values from Whitfield (1957). Measurements are difficult and we do not know if the extreme value for NGC 6166 has been confirmed.

(9) From Mills (1959 d).

(10) From Mills (1959 d), excepting 6166, Ryle (1958). Various authorities quote widely divergent values depending on whether bright parts or extreme outer limits are specified. The figures quoted tend to be the higher of commonly quoted values, but are probably somewhat inconsistent.

(13) Based on columns (4) and (9).

(15) P in w (c/s)$^{-1}$ sterad^{-1} at 3·5 m defined by $S = P/r^3$; distances from column (4) and S from Mills (1959 d). Note that luminance integrated over the range 10–10^4 Mc/s $\simeq 10^8$ P, and the total power emitted is 4π times more.

(16) Absolute photographic magnitudes; the first six are from Dewhirst (1959) but, for consistency, have been adjusted for a Hubble constant of 180 km sec^{-1} (Mpc)$^{-1}$, (instead of Dewhirst's value of 100). Dewhirst's values are 1·2 magnitudes brighter. Cen A and For A values are from columns (4) and (6).

that all radio galaxies are giant galaxies with absolute photographic magnitude $M_p = -20{\cdot}5$.

The statistical distribution of $m_{1\cdot9} - m_p$ followed directly from observational material; it resulted from a search for coincidence between radio source and galaxy, starting from the radio position. Coincidences in position within the observational uncertainty were noted as 'identifications' if the galaxy, if single, was brighter than magnitude $m_p = 16$, or if it was one of a pair or small group, $m_p = 19$. In consequence, as noted by Mills, the derived distribution may be subject to selection effects. An identification would not be recognized with a very distant galaxy of high radio emission, nor with an object of low optical luminosity, and an extension of the distribution on the high radio luminosity side is quite possible.

In connection with Table 5 it was noted that the sizes of radio galaxies greatly exceeded the optical sizes. In two cases, Centaurus A and Cygnus A, the brightness distribution has been studied in some detail. The results for the former, the nearest radio galaxy, have been collected by Wade (1959 b) (see also Bolton and Clark 1960). The radio source extends over an area of $7° \times 2°$, and has an intense peak of angular size about $7' \times 3'$ coinciding closely in position, and approximately in size, with the bright part of the optical nebula. The main part of the emission

arises in the extended part which itself consists of two extended peaks. The appearance is broadly similar at the three frequencies, 19 Mc/s, 85 Mc/s and 1400 Mc/s ; a difference observed at the lowest frequency is that the contribution of the central peak falls from about 25% to 10% of the whole. This may be due to absorption in HII regions. Radio and optical appearances are shown in Fig. 26 and Fig. 27 (Plate III) respectively.

Surprisingly similar results have been obtained for Cygnus A, the most intense of the radio galaxies, by Jennison (1959) and Twiss, Carter and Little (1960). Over the frequency range 127 to 3000 Mc/s the source shows two main peaks,

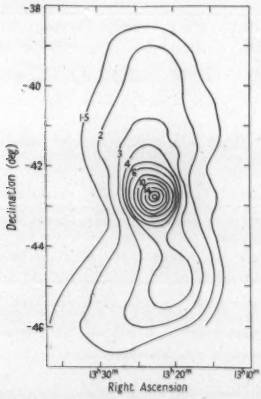


Fig. 26. 85 Mc/s contours of brightness temperature for the radio galaxy, Centaurus A, based on Sheridan (1958) (unit, 3000°K).

separated by about 80″ and symmetrically situated on each side of the optical nebula which is about 18″ × 30″ in extent. The optical object thus lies in the gap between the two radio peaks.

In both cases the main part of the radio emission seems to have very little relation to the optical, and both are elongated objects with a maximum extent of about 100 kpc.

The most extensive investigations on source spectra have been made by Whitfield (1957, 1959), who has collected results from many observers. For the strongest sources the spectra extend from about 20 to 10 000 Mc/s. Neglecting the part near 20 Mc/s which sometimes shows some evidence of a decrease below the simple form, possibly due to HII absorption, the spectra of Class II sources are fitted within the experimental uncertainty by a curve of the form $S \propto f^{-\alpha}$ where α is about 0·7. Data for many weaker sources are available over lesser frequency ranges and it seems that, again within the observational uncertainties which are considerable, all extragalactic sources show such power-law spectra. According to Whitfield

the indices differ somewhat from source to source. Some values are included in Table 5.

The existence of these rare galaxies with exceptionally great radio emission is well established, but we do not understand what leads to this feature. Collisions between galaxies may well be one cause, but as more radio galaxies are identified the percentage of apparent collisions diminishes. Even for those which appear to show collision symptoms, the collisions hypothesis is questioned. Ambartsumian (1958) suggests that these may, instead, be cases of disrupting or exploding galaxies, undergoing fission under the influence of quite unknown forces. Another common feature among radio galaxies is membership of a small group or cluster, which suggests some form of interaction.

The most plausible suggestion as to the mechanism of emission is again the synchrotron mechanism. This is supported by the power-law spectrum, and in a single instance, NGC 4486, by the optical observation of linear polarization in the 'jet' of that peculiar galaxy. Accepting the mechanism, the vast extent of the radio emitting regions both in normal and radio galaxies points to the existence of magnetic fields and high energy electrons far out into intergalactic space. In the vicinity of galaxies what was previously thought to be intergalactic space is not empty. The mechanism, however, raises difficulties concerning the enormous amounts of energy stored in magnetic fields and energetic particles. In our own and other normal galaxies we are familiar with the surprisingly high total energy of cosmic rays; for Cygnus A the energy is many orders higher. For example Burbidge (1959) quotes the following values for the sum of the magnetic and particle energies necessary for the radio emission, on the assumption that these are equal, for Cygnus A and the normal Sb galaxy M 31. He gives alternative values based on the assumption: (i) that the electrons are the primary particles and (ii) that the electrons are secondary particles due to a primary flux of protons.

	Total Energy (Particles and Magnetic) Assuming primary particles are :	
Galaxy	(i) Electrons	(ii) Protons
M 31	$2{\cdot}1 \times 10^{58}$ ergs	3×10^{58} ergs
Cygnus A	$2{\cdot}8 \times 10^{59}$ ergs	$3{\cdot}9 \times 10^{60}$ ergs

For comparison Baade and Minkowski's estimate of the total energy of the postulated Cygnus A collision is 10^{59} ergs. The source of energy of the most intense radio sources is a mystery.

Leaving the question of what causes radio galaxies unanswered, are there any criteria which in the absence of optical identifications should help to sort out the near, intrinsically weak, emitters from the very distant, powerful ones? Table 5 gives radio luminances. It appears that differences in luminosity are primarily associated with differences in surface brightness, not emitting area. Areas, and linear dimensions, tend to be roughly constant. Hence, subject to the limitations of such a small sample, it appears that we may be able to use the angular diameter of a source as a crude measure of its distance, and the brightness (or brightness temperature) as an indication of its intrinsic luminosity.

7.2. *Number Counts, Background Brightness and Cosmology*

The hope raised by the discovery of Cygnus A is that from observations of even more distant radio galaxies we may obtain information of cosmological significance. We are thinking here of effects which are only apparent from observations of extremely distant objects. Such effects may be recognized through their being inconsistent with the simplest ' world model ', an infinite, static, Euclidean universe. But it has been difficult to determine which observations are crucial. Optically, the only two clear-cut examples are the red shift, and the fact that the sky between the stars is dark. On the radio side the two present lines of approach are the darkness of the sky between radio sources, and the statistical intensity distribution of Class II radio sources, assuming these to be radio galaxies.

Consider a region containing a distribution of sources having an average spatial density ρ and suppose each source has a radio luminance P, i.e. it emits a power P watt $(c/s)^{-1}$ sterad^{-1}. Then the number of sources N per unit solid angle having a flux density greater than S is given by

$$N = (\tfrac{1}{3}\rho P^{3/2})\, S^{-3/2} = \beta S^{-3/2}.$$

If the sources are unequal in power a similar proportionality exists in which β is given by $\Sigma\frac{1}{3}\rho P^{3/2}$, summed for a series of groups containing sources having small ranges of P. Thus a flux density distribution in the form of a $-3/2$ power law is the prediction from the static Euclidean model, and the observational deviations from it would be of major importance. It is readily shown that the distribution cannot extend indefinitely because there would not then be dark spaces between the sources. However, the deviation originally put forward by the Cambridge radio astronomers is that the distribution should first show an increase above the $-3/2$ law before this decrease. The current evidence is as follows.

The two most extensive and apparently reliable surveys now available are the Sydney 85 Mc/s Mills Cross pencil beam survey of the southern sky with 1658 sources (Mills, Slee and Hill 1960) and the Cambridge 3C interferometer survey of the northern sky on 159 Mc/s with 242 sources (Edge *et al.* 1959). Both surveys show an approximately isotropic distribution. Figure 28(*a*) shows on logarithmic scales N plotted against S for each of these surveys ; the displacement between the curves is associated with the frequency difference. Both curves are effectively straight lines indicating power laws, but the slopes, which are of special interest, differ. It is difficult to assess the observational uncertainties ; statistical errors are large at the bright end and instrumental effects are possible at the faint end due to operation near the resolution or sensitivity limits of equipment, but the slopes, $-1{\cdot}5$ for Sydney and $-2{\cdot}0$ for Cambridge, appear to differ significantly. Edge *et al.*, from a comparison of individual sources in a region where the two surveys overlap, concluded that the difference arises from the exclusion from the Cambridge survey of sources of large angular size and low brightness temperature which are included in the Sydney survey. The inclusion or exclusion of such sources could selectively affect the statistics and more information about them is required for a proper evaluation of the statistics. In view of the current observational uncertainty, the conclusion to be drawn from these studies is that there is no definite evidence as yet of a real departure from the $-3/2$ power-law distribution.

Further evidence has been put forward by Scheuer (1957) based, not on counts of individual sources, but on the statistics of the occurrence of different amplitudes in interferometer records. The analysis assumes the sources to be randomly distributed in space, and to have small angular dimensions. The predicted distribution of amplitudes disagreed with that observed in the sense to be expected if there were a large excess of faint sources. However, it is quite possible, indeed probable, that the other postulates may not be fulfilled, they could be upset by small-scale clustering or by sources of large angular size, and for this reason we consider that judgement should be reserved on the reason for the discrepancy.

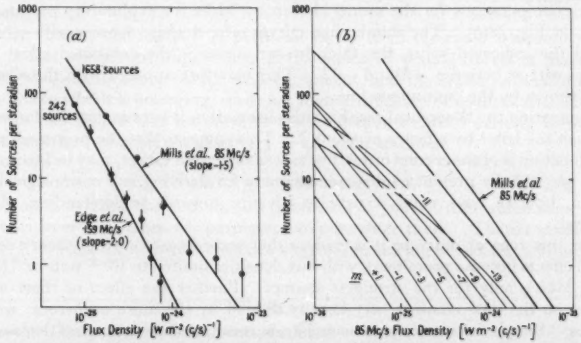


Fig. 28. The intensity distribution of Class II sources; numbers of sources having flux densities greater than assigned values are plotted. (*a*) Observed; (*b*) counts predicted from Mills' luminosity function (thick broken line) are compared with observed counts (continuous line). The other curves give estimated contributions by sources with $\underline{m}$ ($= m_{1\cdot 9} - m_{\mathrm{D}}$) within $\pm$ 1 of the quoted $\underline{m}$.

It is of interest to consider the implications if source counts follow the $-3/2$ power law within the observational limits. The simplest explanation is that, for the bulk of the sources, the source distribution is uniform out to the limit of observation of the bulk of the sources, and that this limit is sufficiently close for cosmological effects to be neglected. On the other hand, as McVittie (1959) has pointed out, the results could equally well conform with a distribution of very intense sources, like Cygnus A, for which cosmological effects (red shift, etc.) severely reduce the flux density from the most distant ones. If so the maintenance of the $-3/2$ law would imply a large increase of either space density or average emission at the greater distances. Further progress then must require further evidence. One possibility is the addition of information on the luminosity function of radio galaxies.

Ryle (1958) has endeavoured to find the typical emission of a radio galaxy from a statistical argument based on (i) flux density—number counts, and (ii) measurements of background brightness (see below) together with an assumed cosmological model. If all sources emitted the same power, (i) gives $\rho P^{3/2}$ and (ii) ρP, so that

together ρ and P would be obtained. On the basis of observations Ryle's argument leads to a value of P of the order of 10^{23} watt $(c/s)^{-1}$ (sterad)$^{-1}$. However, the actual spread in the power emitted by different sources is very great and it is difficult to see how to make the next step.

Mills (1960) has approached the problem from his observed statistical distribution of $m_{1\cdot9} - m_p$ (see Fig. 25). Given the radio luminosity function, the space distribution of galaxies, and the spectrum of the sources (assumed $\propto \lambda$), it is possible to compute, for a given cosmological model and assuming no evolutionary changes in P and ρ, the form of the variation of N with S for sources in each magnitude class, $\underline{m} = m_{1\cdot9} - m_p$, and the integrated counts including all classes. The results of such computations for the model chosen by Mills for exploratory purposes are shown in Fig. 28(*b*). The thin broken curves refer to those sources with $\underline{m}$ within ± 1 of the indicated value, the thick broken curve is the combined effect of all sources with $\underline{m}$ between $+1$ and -13. This may be compared with the observed curve shown by the continuous line.

Comparing the theoretical results with observation it is seen that the former are less than the latter by a factor of about 2. This suggests that the proposed luminosity function is of the correct order, but as was suggested earlier, may be incomplete. Again we find the present evidence inadequate for deriving any cosmological conclusion, but we have made a definite advance towards understanding what is required.

For this type of universe it is evident that according to current theory cosmological effects become prominent, with flux densities limited to 10^{-26} watt m^{-2} $(c/s)^{-1}$ (at 85 Mc/s), only for the strongest sources. Further the effect of these strong sources on the total count is very heavily diluted by the more numerous, weaker, sources. Simple source counts look like a most inefficient way of looking for cosmological effects. The immediate hope is that by selecting sources, say those of high brightness, a sample of much higher luminosity may be achieved which may show cosmological effects. This is one of the motives behind the survey of sources of small angular size in progress at Sydney.

In optical astronomy the blackness of the sky between the stars is the most elementary evidence precluding an infinite, static, Euclidean universe. In the same way the radio sources stand out against a background of much lower brightness, and, if the part of that brightness due to the unresolved faint radio galaxies could be determined, it could supply very important evidence for checking cosmological theories. However, this determination is complicated by the poorly known emission from the corona of our galaxy in the same way that optical observations are upset by the light of the night sky.

Baldwin (1955) and others have made speculative estimates of the residual extragalactic component based on models of the galactic corona. Baldwin finds it to be less than 500°K at 81 Mc/s, i.e. not more than half the brightness in the least bright parts of the sky at this frequency.

Shain (1959), using a lower frequency, has instituted a more definite approach based on the obscuration of the extragalactic background, but not the galactic corona, by an extensive HII region situated outside our galaxy. He used the giant emission region 30 Doradus in the Large Magellanic Cloud. More precise data, both optical and radio, are required for an exact determination but the

available data indicate that at a frequency of 20 Mc/s the contribution of the extragalactic component is less than 22 000°K (and possibly about 7000°K). This corresponds to less than 50% (and possibly about 20%) of the sky brightness in the colder parts of the sky at 20 Mc/s.

The radio evidence on background brightness confirms qualitatively the known inadequacy of the infinite, static, Euclidean universe, but is not yet able to give quantitative values which could provide a real test for the more sophisticated cosmological theories. From source counts there is a suggestion that there may be a discrepancy with the above model, but the current evidence is inadequate to prove it. Radio astronomy has therefore not yet made a positive contribution to cosmology.

What then of the future? A pertinent factor from the radio astronomy viewpoint is that equipment improved in performance by several orders of magnitude should be available soon. With such equipment, and the better understanding of the problem which is emerging, there is a high probability of a substantial advance. Whether this should take the form of discrimination between existing theories or of a completely unexpected discovery is an open question.

Acknowledgments

We wish to thank the authors and publishers of papers from which some of the diagrams have been taken for permission to reproduce them. Thanks are also due to Mr. F. J. Kerr for many constructive suggestions during the writing of this paper.

References

Books. Textbooks on radio astronomy include :

1. Steinberg, J. L., and Lequeux, J., *Radioastronomie* (Paris : Dunod, 1960).
2. Brown, R. Hanbury, and Lovell, A. C. B., *The Exploration of Space by Radio* (London : Chapman and Hall, 1957).
3. Shklovsky, I. S., *Cosmic Radio Waves* (Cambridge, Mass. : Harvard, 1960).
4. Pawsey, J. L., and Bracewell, R. N., *Radio Astronomy* (Oxford : Clarendon, 1955).
5. Smith, F. G., *Radio Astronomy* (Harmondsworth: Penguin Books, 1960).

Of these, 1 is an excellent up-to-date account of all aspects of radio astronomy, 2 is somewhat more introductory, 3 is a translation into English with extensive revision, of the author's Russian text, dated 1956, and gives a particularly good account of Russian work on synchrotron radiation, cosmic rays, etc., 4, although published in 1955, still contains many of the fundamentals of the subject, while 5 (Pelican book number A479) is a recent and authoritative popular account of the subject.

References

Alfvén, H., and Herlofson, N., 1950, *Phys. Rev.*, **78**, 616.
Ambartsumian, V. A., 1958, *La Structure et l'Evolution de l'Universe.* 11th Solvay Conference (Brussels : Stoops) p. 241.
Baade, W., and Minkowski, R., 1954, *Astrophys. J.*, **119**, 206.
Baldwin, J. E., 1955, *Mon. Not. R. Astr. Soc.*, **115**, 684, 690.
Blaauw, A., Gum, C. S., Pawsey, J. L., and Westerhout, G., 1959, *Mon. Not. R. Astr. Soc.*, **119**, 422.
Bok, B. J., 1959, *Observatory*, **79**, 58.
Bolton, J. G., and Clark, B. G., 1960, *Publ. Astr. Soc. Pacif.*, **72**, 29.
Bolton, J. G., and Stanley, G. J., 1948, *Nature, Lond.*, **161**, 312; 1949, *Aust. J. Sci. Res.* A, **2**, 139.

BOLTON, J. G., STANLEY, G. J., and SLEE, O. B., 1949, *Nature, Lond.*, **164**, 101.
BROWN, R. HANBURY, and HAZARD, C., 1951, *Mon. Not. R. Astr. Soc.*, **111**, 357; 1959, *Mon. Not. R. Astr. Soc.*, **119**, 297.
BURBIDGE, G. R., 1958, *Astrophys. J.*, **127**, 48; 1959, *Paris Symp. I.A.U.*, **9**, 541.
BURKE, B. F., 1957, *Astr. J.*, **62**, 90.
CHRISTIANSEN, W. N., 1959, *Proc. Instn Radio Engrs Aust.*, **20**, 519.
CHRISTIANSEN, W. N., and HINDMAN, J. V., 1952, *Aust. J. Sci. Res.* A, **5**, 437.
DEWHIRST, D. W., 1959, *Paris Symp. I.A.U.*, **9**, 507.
DOMBROVSKY, V. A., 1954, *Dokl. Akad. Nauk SSSR.*, **94**, 1021.
DRAKE, F. D., 1959 a, *Paris Symp. I.A.U.*, **9**, 366; 1959 b, *Astr. J.*, **64**, 329.
DRAKE, F. D., and EWEN, H. I., 1958, *Proc. Inst. Radio Engrs N.Y.*, **46**, 53.
DRÖGE, F., and PRIESTER, W., 1956, *Z. Astrophys.*, **40**, 236.
EDGE, D. O., SHAKESHAFT, J. R., MCADAM, W. B., BALDWIN, J. E., and ARCHER, S., 1959, *Mem. R. Astr. Soc.*, **68**, 37.
EWEN, H. I., and PURCELL, E. M., 1951, *Nature, Lond.*, **168**, 356.
GETMANTSEV, G. G., 1952, *Dokl. Akad. Nauk SSSR.*, **83**, 557.
GINZBURG, V. L., 1951, *Dokl. Akad. Nauk SSSR.*, **76**, 377.
GREENSTEIN, J. L., 1957, *Manchester Symp. I.A.U.*, **4**, 179.
GUM, C. S., KERR, F. J., and WESTERHOUT, G., 1960, *Mon. Not. R. Astr. Soc.*, **121**, 132.
GUM, C. S., and PAWSEY, J. L., 1960, *Mon. Not. R. Astr. Soc.*, **121**, 150.
HADDOCK, F. T., MAYER, C. H., and SLOANAKER, R. M., 1954, *Astrophys. J.*, **119**, 456.
HEY, J. S., PARSONS, S. J., and PHILLIPS, J. W., 1946, *Nature, Lond.*, **158**, 234.
HILL, E. R., SLEE, O. B., and MILLS, B. Y., 1958, *Aust. J. Phys.*, **11**, 530.
VAN DE HULST, H. C., RAIMOND, E., and VAN WOERDEN, H., 1957, *Bull. Astr. Insts Netherlds*, **14**, 1; No. 480.
JANSKY, K. G., 1935, *Proc. Inst. Radio Engrs N.Y.*, **23**, 1158.
JENNISON, R. C., 1959, *Paris Symp. I.A.U.*, **9**, 309.
KAHN, F. D., and WOLTJER, L., 1959, *Astrophys. J.*, **130**, 705.
KERR, F. J., and HINDMAN, J. V., 1957, *Publ. Astr. Soc. Pacif.*, **69**, 558.
KERR, F. J., HINDMAN, J. V., and CARPENTER, M. S., 1957, *Nature, Lond.*, **180**, 677.
KERR, F. J., HINDMAN, J. V., and GUM, C. S., 1959, *Aust. J. Phys.*, **12**, 270.
KERR, F. J., HINDMAN, J. V., and ROBINSON, B. J., 1954, *Aust. J. Phys.*, **7**, 297.
KERR, F. J., and DE VAUCOULEURS, G., 1955, *Aust. J. Phys.*, **8**, 508; 1956, *Aust. J. Phys.*, **9**, 90.
KIEPENHEUER, K. O., 1950, *Phys. Rev.*, **79**, 738.
KUZMIN, A. D., and UDALTSOV, V. A., 1959, *Paris Symp. I.A.U.*, **9**, 305.
KWEE, K. K., MULLER, C. A., and WESTERHOUT, G., 1954, *Bull. Astr. Insts Netherlds*, **12**, 211; No. 458.
LARGE, M. I., MATHEWSON, D. S., and HASLAM, C. G. T., 1959, *Nature, Lond.*, **183**, 1250.
LILLEY, A. E., and MCCLAIN, E. F., 1956, *Astrophys. J.*, **123**, 172.
MCCLAIN, E. F., 1960, *Sci. Amer.*, **202**, 45.
MCGEE, R. X., and BOLTON, J. G., 1954, *Nature, Lond.*, **173**, 985.
MCVITTIE, G. C., 1959, *Paris Symp. I.A.U.*, **9**, 533.
MAYER, C. H., MCCULLOUGH, T. P., and SLOANAKER, R. M., 1957, *Astrophys. J.*, **126**, 468.
MAYER, C. H., and SLOANAKER, R. M., 1959, *Astr. J.*, **64**, 339.
MENON, T. K., 1958, *Proc. Inst. Radio Engrs N.Y.*, **46**, 230.
MILLS, B. Y., 1952, *Aust. J. Sci. Res.* A, **5**, 266; 1955, *Aust. J. Phys.*, **8**, 368; 1956, *Observatory*, **76**, 65; 1959 a, *Publ. Astr. Soc. Pacif.*, **71**, 267; 1959 b, *Paris Symp. I.A.U.*, **9**, 431; 1959 c, *Paris Symp. I.A.U.*, **9**, 498; 1959 d, *Handb. d. Phys.*, Ed. S. Flügge, **53**, 239 (Berlin: Springer); 1960, *Aust. J. Phys.*, **13**, 550.
MILLS, B. Y., LITTLE, A. G., and SHERIDAN, K. V., 1956, *Aust. J. Phys.*, **9**, 218.
MILLS, B. Y., SLEE, O. B., and HILL, E. R., 1960, *Aust. J. Phys.*, **13**, 676.
MINKOWSKI, R., 1958, *Rev. Mod. Phys.*, **30**, 1048; 1959, *Paris Symp. I.A.U.*, **9**, 315.
MULLER, C. A., and OORT, J. H., 1951, *Nature, Lond.*, **168**, 357.
MULLER, C. A., and WESTERHOUT, G., 1957, *Bull. Astr. Insts Netherlds*, **13**, 151; No. 475.
MÜNCH, G., 1960, *Astrophys. J.*, **131**, 250.

OORT, J. H., 1959, *Handb. d. Phys.*, Ed. S. Flügge, **53**, 100 (Berlin : Springer).
OORT, J. H., KERR, F. J., and WESTERHOUT, G., 1958, *Mon. Not. R. Astr. Soc.*, **118**, 379.
OORT, J. H., and ROUGOOR, G. W., 1960, *Mon. Not. R. Astr. Soc.* **121**, 171.
OORT, J. H., and WALRAVEN, TH., 1956, *Bull. Astr. Insts Netherlds*, **12**, 285 ; No. 462.
OSTERBROCK, D., and FLATHER, E., 1959, *Astrophys. J.*, **129**, 26.
PAWSEY, J. L., and GARDNER, F. F., 1959, *Proc. Instn Radio Engrs Aust.*, **20**, 528.
PAWSEY, J. L., and HARTING E., 1960, *Aust. J. Phys.*, **13**, 740.
PIKELNER, S. B., 1953, *Dokl. Akad. Nauk SSSR*, **88**, 229.
PIKELNER, S. B., and SHKLOVSKY, I. S., 1957, *Astr. J., Moscow*, **34**, 145 ; 1958, *Rev. Mod. Phys.*, **30**, 935 ; 1959, *Ann. Astrophys.*, **22**, 913.
PURCELL, E. M., and FIELD, G. B., 1956, *Astrophys. J.*, **124**, 542.
RAZIN, V. A., 1958, *Astr. J., Moscow*, **35**, 241.
REBER, G., 1940, *Proc. Inst. Radio Engrs N.Y.*, **28**, 68.
ROUGOOR, G. W., and OORT, J. H., 1960, *Proc. Nat. Acad. Sci., Wash.*, **46**, 1.
RYLE, M., 1958, *Proc. Roy. Soc.* A, **248**, 289.
SCHEUER, P. A. G., 1957, *Proc. Camb. Phil. Soc.*, **53**, 764.
SCHMIDT, M., 1956, *Bull. Astr. Insts Netherlds*, **13**, 15 ; No. 468 ; 1957 a, *Bull. Astr. Insts Netherlds*, **13**, 247 ; No. 475 ; 1957 b, *Bull. Astr. Insts Netherlds*, **14**, 17 ; No. 480.
SHAIN, C. A., 1958, *Proc. Inst. Radio Engrs N.Y.*, **46**, 85 ; 1959, *Paris Symp. I.A.U.*, **9**, 328.
SHAKESHAFT, J. R., RYLE, M., BALDWIN, J. E., ELSMORE, B., and THOMPSON, J. H., 1955, *Mem. R. Astr. Soc.*, **67**, 106.
SHERIDAN, K. V., 1958, *Aust. J. Phys.*, **11**, 400.
SHKLOVSKY, I. S., 1952, *Astr. J., Moscow.*, **29**, 418 ; 1953 a, *Dokl. Akad. Nauk SSSR*, **90**, 983 ; 1953 b, *Dokl. Akad. Nauk SSSR*, **91**, 475 ; 1953 c, *Astr. J., Moscow*, **30**, 15 ; 1954, *Dokl. Akad. Nauk SSSR*, **98**, 353.
SPITZER, L., 1956, *Astrophys. J.*, **124**, 20.
THOMPSON, J. M., 1957, *Nature, Lond.*, **180**, 495.
TUNMER, H., 1959, *Paris Symp. I.A.U.*, **9**, 571.
TWISS, R. Q., CARTER, A. W. L., and LITTLE, A. G., 1960, *Observatory*, **80**, 153.
VOLDERS, L., and VAN DE HULST, H. C., 1959, *Paris Symp. I.A.U.*, **9**, 423.
WADE, C. M., 1958, *Aust. J. Phys.*, **11**, 388 ; 1959 a, *Aust. J. Phys.*, **12**, 418 ; 1959 b, *Aust. J. Phys.*, **12**, 471.
WESTERHOUT, G., 1957, *Bull. Astr. Insts Netherlds*, **13**, 201 ; No. 475 ; 1958, *Bull. Astr. Insts Netherlds*, **14**, 215 ; No. 488.
WHITFIELD, G. R., 1957, *Mon. Not. R. Astr. Soc.*, **117**, 680 ; 1959, *Paris Symp. I.A.U.*, **9**, 297.
WILD, J. P., 1952, *Astrophys. J.*, **115**, 206.
VAN WOERDEN, H., ROUGOOR, G. W., and OORT, J. H., 1957, *C.R. Acad. Sci., Paris*, **244**, 1691.
WOLTJER, L., 1957, *Bull. Astr. Insts Netherlds*, **13**, 301 ; No. 478 ; 1958, *Bull. Astr. Insts Netherlds*, **14**, 39 ; No. 483.

MAGNETIC DOMAINS

By D. J. CRAIK† and R. S. TEBBLE‡

† Boots Pure Drug Company, Nottingham

‡ Department of Physics, Sheffield University

CONTENTS

	PAGE
§ 1. Introduction	117
§ 2. General principles	117
2.1. Free pole effects (magnetostatic energy; shape anisotropy)	118
2.2. Magneto-crystalline anisotropy	119
2.3. Magnetostriction	119
2.4. Exchange energy	120
2.5. Boundary energy and thickness	120
2.6. Closure domains and subsidiary domain structure	122
2.7. The width of domains	123
2.8. The μ^* effect	123
2.9. Single domain particles	123
§ 3. Methods of observation	124
3.1. Bitter figure technique	125
3.2. Modification for electron microscopy	127
3.3. Transmission electron microscopy	127
3.4. Electron beam methods	128
3.5. Electron mirror microscopy	129
3.6. The photoelectron method	129
3.7. Probe method	129
3.8. Kerr and Faraday effects	129
3.9. Comparison of the methods	131
§ 4. Domain structures of materials with uniaxial anisotropy	132
4.1. Domain structures of uniaxial crystals	132
4.2. Materials with low saturation magnetization—single crystals	132
4.3. Materials with low magnetization—polycrystalline specimens	135
4.4. Crystals with high saturation magnetization	136
4.5. The magnetization of uniaxial crystals	137
§ 5. Domain structure in iron	139
5.1. Iron crystals	139
5.2. Whiskers	142
5.3. Polycrystalline specimens	143
§ 6. Domain structure of nickel-like materials	145
6.1. Nickel and face-centred cubic alloys (single crystals)	145
6.2. Domain structure of cubic ferrites—single crystals	147
6.3. Polycrystalline cubic ferrites	149
6.4. Rare earth garnets and ortho-ferrites	151
§ 7. Alloys which respond to magnetic annealing, and powders	153
7.1. Annealed nickel–cobalt	153
7.2. Perminvar	154
7.3. Cobalt–nickel alloys	155
7.4. Precipitation hardened alloys	156
7.5. Powder or elongated single domain magnets	157
§ 8. Thin films	158
§ 9. Domain walls	161
§ 10. Summary	162
Acknowledgments	164
References	164

Vol. 24
1961

Abstract. The paper gives a survey of developments over the last decade in the methods of investigation into the structure of ferromagnetic domains and the authors discuss the results in relation to present-day theoretical ideas on domain structure. Examples of the application of methods using Bitter patterns, electron microscopy and magneto-optical effects are given for a wide variety of materials ranging from metallic single crystals to polycrystalline ferrites.

§ 1. Introduction

This report is intended to be a survey of the methods of observation of ferromagnetic domains and some of the principal results obtained, and as an appraisal of their significance in relation to present-day ideas on the magnetization process. The background to this work—the theory of coercivity, domain formation and hysteresis generally—has already been adequately treated in the *Reports* by Stoner (1948, 1950), Kittel (1949), Kittel and Galt (1956) and Elschner and Andrä (1955), and therefore emphasis is placed on later experimental results not reported in the earlier surveys, in general, contributions published since 1950.

In § 2 are summarized the basic properties of ferromagnetics and the simpler forms of domain structure necessary for development of the later sections of experimental techniques (§ 3) and methods of observation. The results of these investigations are then discussed in the context of a more detailed examination of domain structure, domain nucleation and magnetization processes.

§ 2. General Principles

The necessity for postulating some kind of sub-division of a ferromagnetic body into domains arises from the fact that, although the exchange forces responsible for the parallel alignment of the electron spins are equivalent to internal magnetic fields of the order 10^7 Oe, it is possible to produce large changes in bulk magnetization by relatively small changes in the external applied field, often of the order of 1 Oe. With each domain magnetized to saturation, the change in magnetization takes place by the growth of the more favourably magnetized domains at the expense of those less favourably directed, and with relatively little change in the energy of the system (compared with exchange energy). In view of their experimental demonstration by a variety of methods the existence of domains will be assumed and the problems concerning their formation will be referred to later.

Magneto-crystalline anisotropy forces tend to align the domain magnetization in one of a number of preferred directions (e.g. in iron, parallel to the cube edges of the crystal lattice). Moreover these domains must form so that the effects of demagnetizing fields arising from the free poles are small; these are produced at surfaces where there is a change in the normal component of the magnetization I, i.e. where $\operatorname{div} I \neq 0$, either at the surface of a specimen or at a boundary between adjacent domains. A domain structure satisfying these conditions with the overall magnetization zero is illustrated in Fig. 1(*a*). Further reduction in the energy of the system may be brought about by further sub-division of the structure with resulting alternation in sign of the surface free pole as in Figs 1(*b*) and 1(*c*). The formation of closure domains near the surface of the specimen reduces the formation

of free pole still further (Fig. 1(*d*)). The extent of the sub-division is limited by the presence of boundary walls between adjacent domains, with which there is associated a finite energy (the 'surface tension' effect) considered below. Equilibrium is reached when the reduction in magnetostatic energy brought about by any further sub-division is offset by an increase in boundary energy associated with the increased wall area. A number of factors are thus involved in the consideration of any domain formation.

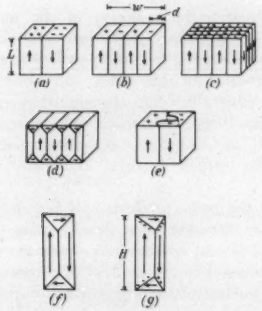


Fig. 1. Examples of domain structure. In the treatment given in the text the depth in (*a*), (*b*) and (*c*) is taken as $L = qw$, with the cross section square of side w.

2.1. *Free Pole Effects* (*Magnetostatic Energy ; Shape Anisotropy*)

When a free pole is produced at a surface lying at right angles to the magnetization vector as at the ends of the specimen in Fig. 1(*a*), the localized field near the surface is $2\pi I_s$ and the energy density, $\frac{1}{2}IH$, is πI_s^2. For ellipsoids of revolution the energy expression can be written $E_D = \frac{1}{2}NI_s^2$ per unit volume, where N is the demagnetizing factor, but in general the energy expression will be a complicated function of I_s^2 and of the geometry of the domain (Stoner 1945, Osborn 1945) ; for this reason the effect is often termed shape anisotropy. The problem of the demagnetizing energy of a rectangular block sub-divided into domains as in Fig. 1(*b*), with each domain uniformly magnetized, was considered, *inter alia*, by Rhodes and Rowlands (1954). In practice the magnetization would not be uniform, but the results would be applicable to a material with high magnetic anisotropy where uniformity of magnetization is nearly complete. It is assumed that there is no contribution from boundary energy.

The demagnetizing energy of a long rectangular block partitioned as in Fig. 1(*b*) is shown in Fig. 2. It can be seen that the reduction in energy with increase in the number of partitions may be considerable but the sub-division is limited by the increase in the boundary energy. When the domain width $d = w/N$ is small (N large), the demagnetizing energy is given by

$$E_D = 1{\cdot}705\, I_s^2\, d\,; \qquad \ldots\ldots(2.1)$$

for the checkerboard structure of Fig. 1(*c*), the corresponding expression given by Kittel (1949) is

$$E_D = 1{\cdot}06\, I_s^2\, d. \qquad \ldots\ldots(2.2)$$

Values of I_s at 20°c in e.m.u. g^{-1} are for iron, nickel and cobalt, 1714, 484 and 1422, respectively.

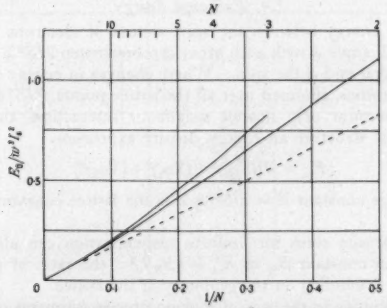


Fig. 2. Reduced demagnetizing energy $E_D/w^3 I_s^2$ of a rectangular block, depth $L = qw$, and square cross section, side w, plotted as a function of N, the number of domains into which the block is divided (Fig. 1(*b*)). The upper curve is for odd and the lower is for even values of N. The broken line is for the limiting expression for $1/N \to 0$, given in Eqn (2.1) (Rhodes and Rowlands 1954).

2.2. *Magneto-crystalline Anisotropy*

The anisotropy energy may be expressed as $\Sigma_n K_n \sin^{2n}\theta$, and for a uniaxial crystal this may be taken as

$$E_K = K_1 \sin^2\theta + K_2 \sin^4\theta \qquad \ldots\ldots(2.3)$$

where θ is the angle between the magnetization and the easy axis, and K_1 and K_2 are the constants typical of the material. Similar expressions apply to cubic crystals, the single angle being replaced by the direction cosines. Since the anisotropy coefficient has the units of energy it can be 'equated' to the magnetic energy term $\frac{1}{2}HI_s$; the equivalent value in terms of magnetic field is thus $H_K = 2K/I_s$. Values of K_1 (and H_K) are, in 10^5 erg cm^{-3} (and Oe): nickel 0·51, (200); iron 4·7, (500); cobalt 43, (6000); MnBi 89, (40 000); the easy directions are: in nickel the cube diagonals, in cobalt and MnBi the hexagonal axis. For cobalt $K_2 = 12 \times 10^5$ erg cm^{-3} (Bozorth 1954).

2.3. *Magnetostriction*

Ferromagnetic crystals generally exhibit a change in dimensions on magnetization (magnetostriction), and when a domain is prevented from expanding freely, e.g. by surrounding domains which are not magnetized parallel to the same direction, a state of strain is created. The corresponding energy density is $\lambda_s \sigma \sin^2\theta$, where λ_s is the saturation magnetostriction (assumed here to be isotropic), σ is the stress,

and θ is the angle between the stress and the magnetization. This magneto-elastic energy may be considered as an addition to the crystal anisotropy, K_1 being replaced by $K_1 + \frac{3}{2}\lambda_s\sigma$. Values of λ_s are: iron -7×10^{-6}, nickel -34×10^{-6}, cobalt -50×10^{-6}, but these isotropic values should be used with caution as λ_s is critically dependent on the direction of magnetization with respect to the crystal axes.

2.4. *Exchange Energy*

The exchange energy between the spin vectors of electrons in neighbouring atoms set at a small angle ϕ with each other can be written $2JS^2\Sigma\cos\phi_{ij}$ where J is the exchange integral and S the spin. Where changes in energy only are involved this expression becomes, summed over all the lattice points, $2JS^2\phi_{ij}^2/2 = JS^2\phi_{ij}^2$.

Taking into account only nearest neighbour interaction, this gives for the body-centred cubic structure an energy density expression

$$E_{ex} = A[(\nabla\alpha_1)^2 + (\nabla\alpha_2)^2 + (\nabla\alpha_3)^2] \qquad \ldots\ldots(2.4)$$

where the exchange constant $A = 2JS^2/a$, a is the lattice constant and α_1, etc. the direction cosines.

This energy density term for uniform magnetization can also be written in terms of the Weiss constant N_W as $E_{ex} = \frac{1}{2}N_W I_0^2$; the ratio of $(A/a^2)/(\frac{1}{2}NI_0^2)$ is of the order 10^{-1}, depending on the geometry of the lattice.

There are difficulties in the way of making precise estimates of the value of A (see Stoner and Wohlfarth 1948, Kittel 1949), but the values given by Stoner (1950) for iron, nickel and cobalt are, in 10^{-6} erg cm^{-1}, 0·83, 0·34 and 1·03.

2.5. *Boundary Energy and Thickness*

The presence of a transition layer or boundary wall between domains involves an increase in the energy of the system. The exchange forces tend to bring about parallel alignment of the spin vectors, and the rotation of the direction of magnetization through the thickness of the wall must therefore be associated with an energy expression which can be derived from Eqn (2.4). There is, in addition, the effect of the magneto-crystalline anisotropy in opposing the rotation of the vectors from the easy direction and thus tending to restrict the width of the boundary. In view of the importance of these factors in the following sections, it is useful to consider the simple example, following Stoner (1950), of a 180° boundary in a uniaxial material such as cobalt (Fig. 3).

The boundary energy term is then

$$E_\gamma = \int [A(\partial\phi/\partial x)^2 + K\sin^2\phi]\,dx \qquad \ldots\ldots(2.5)$$

where A is the exchange constant, K the first anisotropy coefficient, ϕ the angle between the magnetization vector and easy direction and x the distance from the centre of the boundary (Fig. 3). This energy expression is a minimum when the two terms in (2.5) are equal, i.e. when

$$A(d\phi/dx)^2 = K\sin^2\phi, \quad \text{with} \quad \cot\phi = -\sinh(x/\delta_0).$$

$\delta_0 = (A/K)^{1/2}$ is a reduced unit of length and $\gamma_0 = (AK)^{1/2}$ is a convenient parameter of energy.

On substituting for ϕ in the original energy equation (2.5) above, expressions for the boundary energy (per unit area) γ and for an ' asymptotic ' boundary width δ are obtained :

$$\gamma = 4(AK)^{1/2} = 4\gamma_0 \quad \text{and} \quad \delta = \pi\delta_0 = \pi(A/K)^{1/2}.$$

The order of magnitude for γ is 1 erg cm^{-2} and for δ is 10^3 Å.

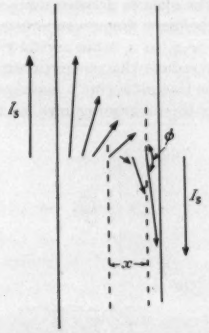


Fig. 3. Direction of magnetization I_s in a domain boundary. ϕ is the angle, at a distance x from the centre of the boundary, between I_s in the boundary and the easy direction in the domain.

Table 1. Effective Domain Boundary Widths δ and Boundary Energies γ given in Reduced Terms δ/δ_0 and γ/γ_0

$\delta_0 = (A/K)^{1/2}$, given in Å ; $\gamma_0 = (AK)^{1/2}$, given in erg cm^{-2}. The domains are magnetized in one or other of the easy directions of magnetization ; the angle between domain magnetization on each side of the boundary is given as ' type of boundary ' (after Lilley 1950).

m.s.—corrected for magnetostriction.

		Direction of normal to domain boundary							
	Type of boundary	[00$\bar{1}$]	[$\bar{1}$10]	[111]	[112]	[001]	[110]	[111]	[$\bar{1}\bar{1}$2]
		Reduced width (δ/δ_0)				Reduced energy (γ/γ_0)			
Nickel	70·53°	3·85	4·26	—		0·54	0·46	—	—
$\delta_0 = 260$;	109·47°	∞	3·31	3·85		1·09	1·37	1·29	
$\gamma_0 = 0{\cdot}14$	180°	—	∞	—	7·92	—	1·83	—	2·00
	180° (m.s.)	—	7·91	—	4·45	—	2·19	—	2·27
Iron	90°	3·14	3·97	3·14		1·00	1·73	1·19	
$\delta_0 = 130$;	180°	∞	5·60			2·00	2·76	—	
$\gamma_0 = 0{\cdot}62$	180° (m.s.)	10·87	5·59			2·02	2·77	—	
Cobalt	180°		3·14				4·00		
$\delta_0 = 50$;									
$\gamma_0 = 2{\cdot}05$									

A detailed treatment of this problem is given by Lilley (1950), who calculates the energy and thickness of boundary walls for cubic and uniaxial crystals, together with corrections for magnetostriction. These are summarized in Table 1.

2.6. *Closure Domains and Subsidiary Domain Structure*

As already mentioned, the main domain structure of a crystal may be modified by the formation of closure domains, in iron-like crystals, of the form shown in Fig. 1(*f*). Only at the corners would there be any free pole in a demagnetized specimen ; for a finite magnetization the picture would take the form shown, with the free pole spread out over the closure domain walls so as to reduce the magnetostatic energy (Fig. 1(*g*)). Subsidiary domain structures also arise at inclusions and imperfections in the crystal ; e.g. in a cubic crystal the well-known Néel ' spike ' structure (Fig. 4(*a*)). These reduce the magnetostatic energy by spreading over the subsidiary domain walls the free pole which would otherwise form on the surface of the inclusion (in some cases the divergence may be abolished altogether).

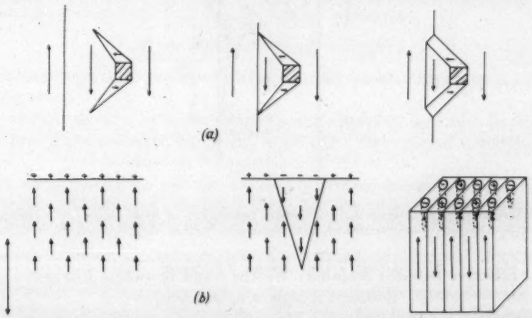


Fig. 4. (*a*) Formation of closure domains (Néel spikes) at an inclusion ; when a 180° boundary wall passes near the inclusion the structure is modified as shown. (*b*) Formation of domains of reverse magnetization at the surface (basal plane) of a uniaxial crystal.

In a material with a single easy direction of magnetization, such as uniaxial barium ferrite, the closure domain structures would involve considerable anisotropy energy, and simple ' spikes ' or ' daggers ' of reverse magnetization at the surface, as shown in Fig. 4(*b*), are more likely to occur, giving an alternation of pole sign in the basal plane with consequent reduction in energy.

The growth of these subsidiary domains of reverse magnetization has been considered by Goodenough (1954), and it is clear that this process will normally take place in high fields as the first stage in the production of full sized domains. The mechanism of the nucleation of these ' spikes ' is still the subject of theoretical study. The energy required for their formation (and destruction) is high, and it is possible that even in extremely strong fields they are not destroyed but remain as ' nuclei ' to redevelop when the field is reduced. The nuclei may be formed initially at flaws in the structure during cooling through the Curie point, where the energy of formation required would be much lower. (This, and similar points, will be discussed in the succeeding sections.)

The picture of magnetization would thus involve in low fields the movement of the 180° boundary across a domain and its impedance at imperfections in the crystal,

with or without the formation of subsidiary domains, and changes in the magnetostatic energy; in increasingly high fields a diminution in size of the spikes of reverse magnetization, followed by their subsequent growth and expansion as the field is reduced.

2.7. *The Width of Domains*

The width of domains with boundaries of finite wall energy has been evaluated for a wide variety of domain structures, but it is sufficient to consider as an example the configuration shown in Fig. 1(*b*) for a uniaxial crystal. As stated in § 2.1 (Eqn (2.1)) the magnetostatic energy, when the number of domains N is large and the spacing d is small, can be written as $1{\cdot}705I_s^2 d$ per unit area of cross section; the wall energy is $\gamma L/d$ where L is the length of the specimen.

The total energy is thus

$$f = \gamma L/d + 1{\cdot}7I_s^2 d$$

and is a minimum when

$$d = (\gamma L/1{\cdot}7I_s^2)^{1/2}, \text{ giving } f_{\min} = 2(1{\cdot}7I_s^2\gamma L)^{1/2}. \quad \ldots\ldots(2.6)$$

The corresponding expressions for the 'checkerboard' structure (Fig. 1(*c*)) are (again to a sufficient approximation cf. Eqn (2.2))

$$d = (2\gamma L/I_s^2)^{1/2} \text{ and } f_{\min} = 2(2I_s^2\gamma L)^{1/2}. \quad \ldots\ldots(2.7)$$

With the closure domain of Fig. 1(*d*) the free pole energy is zero, but there is a contribution from the magnetocrystalline energy of $(Kd/2)$ per unit area of cross section and we have

$$d = (2\gamma L/K)^{1/2} \text{ and } f_{\min} = (2\gamma KL)^{1/2}. \quad \ldots\ldots(2.8)$$

The dependence of d upon a term in $L^{1/2}$ is a general characteristic of these expressions.

2.8. *The μ^* Effect*

Domain formation is modified, in particular near the surface of a specimen, by the effect of the demagnetizing fields in producing a rotation of the magnetization vector away from the easy direction. The extent of the rotation depends on the relative values of the free pole energy ($\propto I_s$) and the magneto-crystalline anisotropy ($\propto K$). In a uniaxial crystal, with the direction of magnetization parallel to the easy direction, the magnetization vector will rotate with a permeability $\mu^* = 1 + 2\pi I_s^2/K$. The energy may therefore be reduced either by the formation of reverse 'spikes' on the basal surface or by the rotation of the magnetization vector near the surface.

2.9. *Single Domain Particles*

As the dimensions of the specimen are reduced the relative magnitude of the boundary energy becomes large compared with the magnetostatic energy and the conditions for the formation of domains (and therefore of domain walls) become progressively less favourable until a critical size is reached, of the order of the boundary thickness, below which sub-division into domains is no longer energetically favourable. The coercivity of such a single domain particle is then a measure of the field necessary to rotate the magnetization vector. For non-spherical particles with low crystal anisotropy, this field must be sufficient to turn the magnetization away from the direction where the demagnetizing field is a minimum

(i.e. parallel to the longest dimension). For a prolate ellipsoid (semi-axes a and b) the difference between the demagnetizing field in two directions is $(N_b - N_a)I_s$, where N_b and N_a are the principal demagnetizing factors, and may be as high as $4\pi I_s$, i.e. of the order 10^4 oersteds for iron. For a collection of randomly orientated single domain particles the expression for the coercivity is $H_c = 0{\cdot}479(N_b - N_a)I_s$ (Stoner and Wohlfarth 1948). When the preferred directions of magnetization are decided by crystal orientation rather than shape, the coercivity term becomes $2K/I_s$, analogous to the first of the above ; for strain anisotropy the expression is $3\lambda_s\sigma/I_s$ (cf. §§ 2.2 and 2.3).

Stoner and Wohlfarth (1948) give for the critical diameter of ellipsoids of rotation

$$D = 2b = \frac{2\pi}{I_s}\sqrt{\left(\frac{2A}{N_a}\right)} \qquad \ldots\ldots(2.9)$$

where A is the exchange constant ; it is assumed that $D \ll \delta$, the boundary thickness (in effect that K is small). The value of D for ellipsoids with axial ratio $m = 5$, 10 and 20, calculated from the above expression and for needles of corresponding length to diameter ratio (from Kittel 1946), are given in Table 2. For a material of very high anisotropy (i.e. $\delta \ll D$) Kittel (1949) gives the critical diameter of a sphere as $D = 188/4\pi I_s^2$, giving for MnBi $D \sim 4000$ Å and for cobalt $D \sim 600$ Å ; thus the values for cobalt given in Table 2 do give orders of magnitude, even though the condition $D \ll \delta$ is not satisfied.

Table 2. Values of the Critical Diameter $D = 2b$, in Å, for Ellipsoidal Single Domain Particles calculated from Equation, using the Values of A given in § 2.4

The values for the cube and for needles are those given by Kittel (1949). The condition $D < \delta$ is clearly satisfied only for a limited number of cases.

	$m = 1$	$m = 5$	$m = 10$	$m = 20$
Nickel ($\delta = 2060$)	520	1280	2120	3700
Cobalt ($\delta = 160$)	310	760	1260	2180
Iron ($\delta = 1410$)	230	560	940	1620
Iron (cube and needle)	150 (cube)	280	790	2400

§ 3. Methods of Observation

The ' classical ' method of detecting changes in the magnetization by means of a search coil wound round the specimen and connected to a galvanometer does not give direct evidence of the presence of a domain structure. It is easy however to demonstrate discontinuities in the magnetization process by this method and to deduce the form of the motion of a domain wall. Similarly, it should in theory at least be possible to make similar deductions from measurements of the thermal and dimensional changes which accompany changes in magnetization. These methods would commonly be considered as ' indirect ' whereas the process of observing the domain structure at the surface of the specimen would be looked upon as a ' direct ' method. But it is not possible to examine the interior of the crystal except for certain thin transparent specimens, which tend in any case to have simple structures, and what is generally observed are patterns formed by the

intersection of the domain structure with the surface of the specimen ; some care must therefore be taken in the deductions made from them.

The methods of observing surface domain patterns fall into two groups. The first involves the detection of the localized magnetic fields produced by changes in the component of the magnetization normal to the surface. In the second group it is change in direction, or magnitude, of the magnetization alone which is detected ; as for instance by a rotation of the plane of polarized light.

The preparation of the specimen usually requires considerable care if the domain structure is not to be obscured by the presence of strains and imperfections near the surface. Although there are materials which require little or no preparation, it is usually necessary to subject the surface of the specimen to mechanical polishing and for this to be followed by electro-polishing to remove the strained layer caused by the mechanical preparation ; annealing may also be required. Some special details of surface preparation are mentioned in the next section.

3.1. *Bitter Figure Technique*

This, the most fruitful method of observing domain structure, depends upon the action of the surface magnetic fields in attracting the fine particles of a magnetic powder spread upon the surface of the specimen (Bitter 1932). The force acting

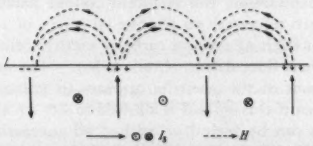


Fig. 5. Representation of the localized fields arising at the intersection of domains with the surface of a specimen. The direction of magnetization in the domains is normal to the plane of the diagram with the vector ' rotating ' through the walls, which are shown for simplicity as single vectors normal to the surface.

upon a small particle volume v, susceptibility κ in a field H, is $\kappa v \operatorname{grad}(H^2)$, and is towards the region of maximum field gradient. Thus in Fig. 5, with the overall domain magnetization parallel to the surface, the powder collects where a domain boundary wall intersects the surface. (It is within the domain wall that the *direction* of magnetization changes producing free pole and localized field gradients at the intersection of wall and surface.)

Should the surface of the specimen be oriented at a small angle with the direction of overall domain magnetization, free poles will be produced, alternating in sign on successive domain surfaces (Fig. 5). The application of an external field normal to the surface will then polarize the colloid particles which collect where the normal component to the magnetization is in the same direction as the applied field, i.e. at alternate domains, and the pattern consists of bands of colloid on the surface.

The colloid will also collect at suitably oriented cracks and flaws in the surface where discontinuity in the magnetization gives rise to free poles ; this is of assistance in finding the direction of magnetization in a domain since colloid is preferentially attracted to those cracks and imperfections lying at right angles to the direction of magnetization.

The ferromagnetic powder is usually prepared, and applied to the surface, as a colloidal suspension of magnetite, Fe_3O_4, in an aqueous medium (although organic liquids and dry powder have been used to extend the range of temperature (Andrä 1956 a, b, 1959)). The first stage of the preparation is to precipitate the magnetite by adding caustic soda (5 g in 50 cm³) to the appropriate mixture of $FeCl_2.4H_2O$ and $FeCl_3.6H_2O$ (2 g and 5·4 g in 300 cm³) and stirring, when particles of colloidal dimensions are produced. The precipitate is then filtered off and washed with distilled water. From this point methods differ.

In Elmore's method (1938 a) the filtrate is washed with 0·01 N hydrochloric acid, then added to a 0·5% solution of soap (sodium oleate) and boiled until a permanent suspension is produced. Using Bozorth's recipe (1951) the filtrate is added to 2 g of dodecylamine in 10 cm³ of N hydrochloric acid. This is diluted to 180 cm³ and stirred by means of a rod rotating at 6000 rev/min, until a colloidal suspension is produced, then further diluted. Both solutions tend to stain the surface of the specimen so that it is sometimes advisable to apply a thin layer of varnish before applying the colloid, provided that this does not affect the definition of the pattern.

A colloid which has little tendency to stain the surface may be produced by dispersing the washed slurry of magnetite in 0·1 N hydrochloric acid. This should cause very effective peptization, the resultant colloid passing easily through fine filter papers. It is then necessary to remove all traces of acid by careful dialysis, and a stabilizing agent such as sodium carboxy methyl cellulose (' Celacol ') must first be added to prevent flocculation (Craik 1956). The same mechanism which prevents the aggregation of the particles appears to prevent their absorption on to the surfaces and even if this colloid is allowed to dry on almost any material but iron the resulting film can be peeled or washed off successfully.

Colloids produced in this way are very well dispersed, and suitable for studying fine detail. They give patterns with relatively low contrast, however, and since the intensity of the patterns depends upon the wall-widths, they are not suitable for the study of materials with low anisotropy (where fine detail cannot, in any case, occur). The use of a true colloid is advantageous when high fields are applied parallel to the surface since a coarse suspension would be swept away.

The method of observing the patterns is straightforward and only involves the application of a small drop of the chosen colloid solution to the surface of the specimen and examination under a microscope. There is some latitude in the choice of density of the colloid and a magnification of a few hundred is usually sufficient for observing the main features of the pattern. Dark-field illumination may be used to increase the contrast, and improved resolution may be obtained by using a water-immersion objective (Moon 1959).

If the surface to be examined is not flat a uniform layer of the colloid is not obtained. In this case, and also for the examination of very small crystals which can be cemented to a microscope slide for the purpose, it may be more convenient to allow a thin layer of Celacol-stabilized colloid to dry on the surface. Very clear patterns may be formed in this way, and can be examined at leisure although continuous wall-movement cannot be followed, and the method cannot be applied to iron (Craik and Griffiths 1958).

3.2. *Modification for Electron Microscopy*

Certain very fine structures are difficult to observe by optical microscopy, since if the colloid is sufficiently dilute it affords very little contrast. To produce patterns suitable for electron microscopy it is necessary to use a colloid which dries to form a thin film without aggregation of the particles, and for this reason the Celacol-stabilized colloid was developed (Craik and Griffiths 1958). The film, in which the particles are embedded in their original distribution, representing the domain structure, is prepared by the usual procedures and strengthened by

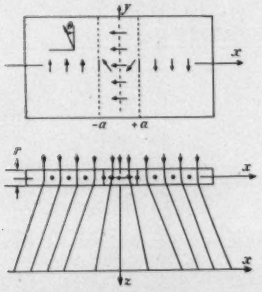


Fig. 7. Deflection of a beam of electrons passing through a domain wall $(-a+a)$ in a thin film (Fuller and Hale 1960).

a thin layer of evaporated carbon. If the surface has special features in relief these are represented by the normal replica effect, and this may be intensified by evaporating the carbon at an angle as exemplified by Fig. 40 (Plate X).

Very low initial magnification is generally adequate, and useful information may be obtained from micrographs at ' optical magnifications ' due to the high contrast. For example, the fine strain patterns on a cobalt crystal, shown in Fig. 21 (Plate III), could in principle be resolved by optical microscopy, but though they must have been present on many of the specimens so examined they were first detected by electron microscopy.

The resolution of the patterns is governed by the nature of the material. When the anisotropy is high the edges of domains in carefully prepared films have a ' spread ' no greater than the diameter of the colloid particles, i.e. the resolution is about 0·02 μ. On crystals with lower anisotropy the edges are more diffuse resulting from the greater wall-width itself and weaker forces on the colloid particles. In some cases the resolution of the individual particles, and their orientation in the stray fields, is of value in the interpretation of the patterns.

3.3. *Transmission Electron Microscopy*

Thin metallic films are significantly transparent to electron beams as used in conventional electron microscopes. By over-focusing the condenser to give a small distant effective source of electrons and using the full magnifying power of the

projector lenses (with the objective inoperative) a low magnification shadow image of the specimen film is obtained in which the domain boundaries are clearly shown (Figs 6 (Plate I) and 7, Fuller and Hale 1960). In the Plate the boundaries are alternately dark and light according to whether the electron trajectories diverge or overlap after passing through the film. It is possible to obtain high-resolution pictures, using the objective lens operated at reduced strength. This elegant method represents a great advance in giving direct measurements of the wall-widths. The results are discussed in the section on thin films, § 8.

The field of the electromagnetic objective lens used in the above work affected the magnetization of the specimen, and electrostatic lenses are to be used to avoid this effect (Fuller, private communication).

3.4. *Electron Beam Methods*

Since the early work of Germer (1942) there have been a number of developments in the techniques which have simplified considerably the analysis of the results. In the method of Blackman and his co-workers a beam of electrons

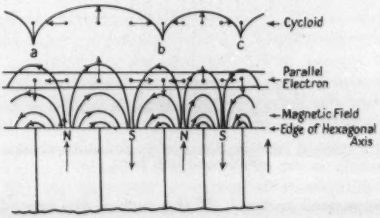


Fig. 8. Diagram showing the deflection of a parallel beam of electrons due to the magnetic field near the surface of a crystal. The source of electrons is above the plane of the diagram (Blackman and Grünbaum 1957).

(~ 50 kev) from an aperture impinges on the crystal where it is diffracted or scattered either at the crystal surface or edge. The electron beam can be produced either as a parallel beam limited by an aperture of 6–25 μ diameter ; or by focusing the original electron beam on the aperture, as an electron beam diverging in ' all ' directions to produce a ' shadow ' photograph of the edge of the specimen. The effect on a parallel beam passing close to the edge of a hexagonal face (perpendicular to the hexagonal axis) is illustrated in Fig. 8. Near the surface the lines of force are as shown and the individual electrons are deflected in the direction shown by the arrows. The beam is distorted into a trochoidal form, passing from a loop pattern to a cycloid and finally to an undulatory pattern with increasing distance from the surface of the crystal, as illustrated in Fig. 9(*a*) (Plate II). The dimensions of the trochoid permit calculations of the strength of the magnetic field acting on the electrons (proportional to the amplitude of the curve) and the domain spacing (proportional to the distance between successive cusps or loops) (Blackman and Grünbaum 1957, 1958).

A typical shadow photograph is shown in Fig. 9(*b*) ; the shadow edge is distorted, again into the form of a cycloidal curve.

PLATE I

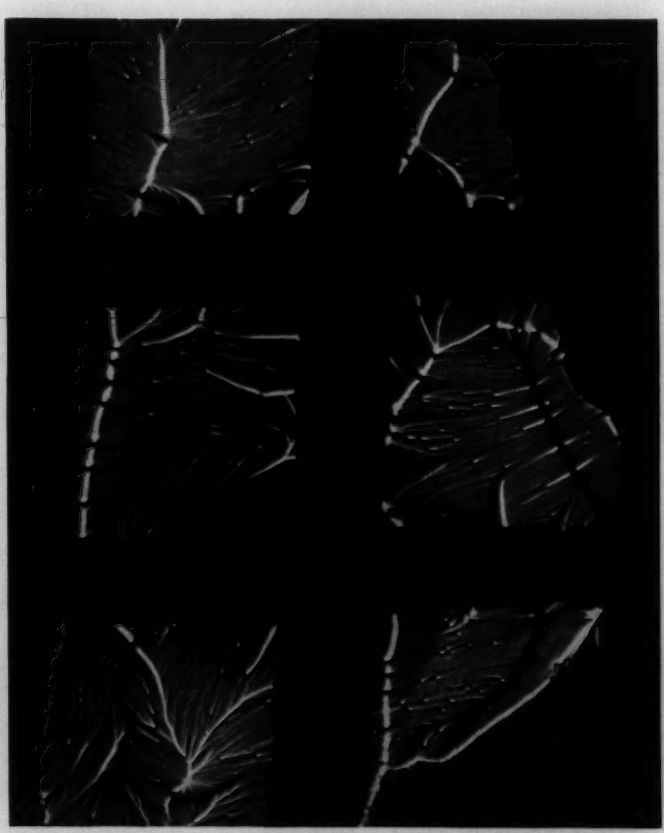


Fig. 6. Transmission electron-micrograph of domain walls (grid 200/in.) (Fuller and Hale 1960).

PLATE II

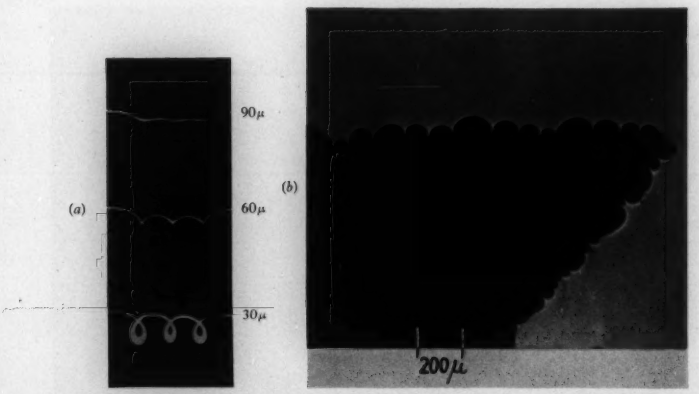


Fig. 9. (*a*) Trochoid patterns produced by an electron beam passing close to the edge of a hexagonal face of a cobalt crystal (distance from edge in microns). (*b*) Typical electron beam shadow photograph (Blackman and Grünbaum 1957).

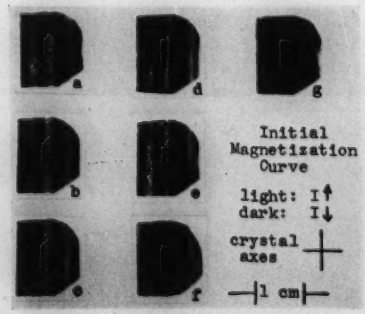


Fig. 11. Domain patterns on the surface of a crystal of cobalt observed by the Kerr effect (Fowler and Fryer 1954 a).

Fig. 15. Domain pattern on grain-oriented barium ferrite.

In the method of Marton, Simpson and Lachenbruch (1954) the domain field (due to the specimen) under study is placed in the path of an electron beam, in the object space of an electron-optical lens system whose image space contains a wire grid. The distortion of the shadow of this grid due to the presence of the domain field can then be measured as a function of position. From the shadow distortion and the geometrical parameters of the electron-optical system, the angular deflections of the different rays caused by the field are calculated with some difficulty.

3.5. *Electron Mirror Microscopy*

In this method a beam of electrons is directed on to the surface of the specimen which is held at a negative potential with respect to the cathode of the electron gun. In this way the electron beam is reflected from a potential surface a short distance in front of the specimen which thus acts as an electron mirror. The image, formed on a suitably placed screen by the reflected electron beam, reproduces the irregularities in the reflecting potential surface caused by the detailed structure of the surface of the specimens. The presence of a domain structure near the surface of the specimen is also detected. The resolution is as good as that obtainable by the usual optical methods but has the advantage that the temperatures of the specimen can be changed and the corresponding alterations in the domain pattern can be followed (Spivak 1957, Mayer 1957).

3.6. *The Photoelectron Method*

The surface of the domain is here coated with a material such as SbCs which acts as a source of photoelectrons when illuminated by a suitable source of light. The electrons emitted are deflected by the magnetic field of the domain structure and after passing through a suitable electron lens system reproduce the domain pattern on a fluorescent screen.

3.7. *Probe Method*

A fine rod of Permalloy is caused to vibrate and placed with its end near to the surface of the ferromagnetic crystal. The presence of a magnetic field is detected by the presence of an alternating potential induced in the probe so that by moving such a rod across the surface of the specimen the stray fields can be measured and the domain configuration mapped out (Kaczér 1955). Although the method allows considerable freedom in the manipulation of the specimen the resolution is limited by the diameter of the probe, about 10^{-2} cm, and it has not been used to any considerable extent.

3.8. *Kerr and Faraday Effects*

The presence of magnetic domains can also be detected by their effect upon the reflection of plane polarized light incident on the domain surface; the reflected light is elliptically polarized, and this, to a sufficient approximation, can be considered as a rotation of the plane of polarization, proportional to the intensity of magnetization in the surface. The incident light should be polarized either parallel or perpendicular to the plane of incidence, to avoid the elliptic polarization which would otherwise be produced by reflection from a metallic surface. The

effect of the relative orientation of the plane of incidence with respect to the direction of magnetization is shown in Fig. 10. The magnitude of the polar effect is several times greater than that of the others but even then it can only be applied to materials of high uniaxial anisotropy when the magnetization vector lies at right angles to the surface.

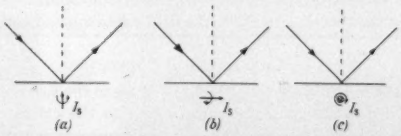


Fig. 10. Kerr effect: the plane of incidence of the plane polarized light is parallel to that of the diagram. (*a*) Polar effect: $\mathbf{I}_s$ perpendicular to the reflecting surface; (*b*) longitudinal effect: $\mathbf{I}_s$ in the plane of the reflection surface and parallel to the plane of incidence; (*c*) transverse effect: $\mathbf{I}_s$ in the plane of the reflecting surface and perpendicular to the plane of incidence (Lee, Callaby and Lynch 1958).

The method consists essentially of illuminating the surface of the ferromagnetic crystal with light, plane polarized by means of a Nicol prism, and viewing the reflected light through a second prism, crossed to extinguish the light in the absence of a domain structure. The presence of domains is shown by the production of light and dark bands when the prism is rotated through the appropriate angle. The method is not however straightforward; Fowler and Fryer (1954 a) found it necessary to develop a special technique to eliminate the 'background' due to surface imperfections. A transparent positive was made of the specimen in the magnetically saturated state, i.e. with little or no domain structure, and this was superimposed upon a negative taken with the required domain pattern present on the crystal surface. The photograph shown in Fig. 11 (Plate II) was obtained by this method. It was suggested by Noskov and Sokolov (1947) that 'blooming' the specimen surface should lead to improved contrast, and this was found to be so when the process was applied to thin films (Prutton 1959). Kirenski and Degtiarev (1959) noted that the contrast in patterns obtained on crystals of silicon iron was considerably improved if a thin layer of iron oxide was formed on the surface, obviating the need for the double photographic process.

As a result of the difficulties which arise in the use of polarized light attempts have been made to improve on the method of detection by the use of a photomultiplier in place of a photographic plate. In the method of Lee, Callaby and Lynch (1958) a small area of the specimen near a domain boundary is illuminated and the reflected light is received on the cathode of a photomultiplier. The movement of the domain boundary is made periodic by the application of a small alternating field, so that the motion of the wall can be followed across the illuminated surface as a change in intensity which is detected by the photomultiplier as a change in amplitude in an a.c. signal.

Results have also been obtained by the transmission of polarized light using the Faraday effect, but here the specimen must be thin enough to be transparent and this places a definite limitation on the nature of the investigation. Results so obtained on thin crystals of barium ferrite and garnets are described in §4.

3.9. *Comparison of the Methods*

The Bitter figure technique has given by far the most results up to the present, and remains the simplest so far as equipment is concerned. With the electron microscope modification the resolution is extended to a fraction of a micron, i.e. to the order of the wall-widths so that any structure should be capable of resolution.

There are two major limitations to the Bitter figure technique. First, the formation of the patterns depends on the gradient of the local fields over the walls and thus on the crystal anisotropy, and patterns fail to form as the anisotropy falls towards 10 erg cm^{-3}. Second, the method can be used over limited temperature ranges only, although this can be increased by modifying the composition of the colloid: suspensions in liquids with low vapour pressure can be used above 100°c, and suspensions in liquid oxygen have been prepared for low temperature studies (Martin, private communication).

The methods utilizing the Kerr and Faraday effects are far less dependent on temperature, and observations have been made on silicon–iron heated to 700°c, despite the fact that the specimen was glowing brightly at this stage (Kirenski and Degtiarev 1959). Moreover these methods depend on the direction of the magnetization itself, and not on the stray fields above the surface, and so are not subject to the requirement of high anisotropy. Although the patterns obtained by the Faraday effect can be produced with the full resolution of the optical microscope the application is limited to crystals which are thin enough to be transparent (about 1 μ for barium ferrite, and a little more for the garnets), while the patterns obtained by the Kerr effect have poor resolution. In recent results on iron whiskers, a resolution of a few microns was obtained by paying particular attention to the angle of incidence of the light beam and the aperture of the lens used (Fowler, Fryer and Treves 1960), and these patterns seem to be the clearest so far produced (§4). The polarization methods are particularly suitable for dynamic studies, and it is possible to observe the effects of a rapidly alternating field on the structures.

The electron scattering experiments, while not generally giving a direct indication of the domain structures, are the most suitable for studying the temperature dependence of the stray fields. The specimens are in a good vacuum, and could presumably be heated up to the Curie points with little difficulty. The electron mirror microscope also is suitable for experiments in which the temperature of the specimen is adjusted, and has the advantage of giving a direct image of the domain structure, though not a measure of the stray fields, but these methods pick out surface imperfections as well as domain features.

The application of transmission electron microscopy to the study of thin films (Fuller and Hale 1960) is clearly important as the full resolution of the electron microscope is utilized; although this should only be necessary for studying the fine structure and width of the walls themselves. It would be useful if such methods could be applied to specimens of substantial size, and it is suggested that reflection electron microscopy (e.g. Oatley and Everhart 1957) should at least be worth investigation. Since the angle of incidence of the electrons can be made small, the path length in any stray fields near the specimen surface may be sufficient for a magnetic pattern to be superimposed on the representation of the surface. It must be remembered however that the resolution of the reflection electron

microscope falls far short of that of the instrument used for transmission. Another type of experiment which might yield useful information would consist of recording the scattering of relatively slow electrons at the edges of fine particles supported on a film in the usual way. This could conceivably give results on the structure of very fine isolated particles which cannot be obtained in any other way. Some of the problems involved in the investigation of domains have been considered in a survey by Bates (1959).

§ 4. Domain Structures of Materials with Uniaxial Anisotropy

The domain structure of any particular crystal or specimen depends on its shape and the orientation of its surfaces with respect to the crystal axes, as well as on its magnetic and other physical properties and, in some cases, on its magnetic history. Thus, rather than attempting to elaborate in a general way on the principles laid down in the Introduction, the structures which have been satisfactorily determined in specimens of various types will be classified so as to indicate the range of existing knowledge, and particular aspects of these results will be discussed later. The structures found in uniaxial materials are most easily described, due to the simplicity of the basic structures and the fact that they are less dependent on the specimen shape and surface orientation than materials with cubic crystal structure.

Unless the contrary is stated the structures described will be those corresponding to the demagnetized state, the behaviour of the structures on magnetization being described separately.

4.1. *Domain Structures of Uniaxial Crystals*

The common uniaxial crystals all possess a high crystal anisotropy (all of order 10^6 erg cm^{-3}), generally referred to the principal axis of a hexagonal crystal structure. The basic structures are arrays of 180° domains, with variations in the subsidiary structures formed at surfaces normal to the easy direction. The formation of closure domains results in a high contribution to the energy from anisotropy effects (proportional to K) whereas a free pole structure involves magnetostatic energy (proportional to I_s^2). Since the anisotropy is approximately of the same order the materials may be conveniently considered in two groups according to the magnitude of their saturation magnetization. The first group, with low I_s, comprises barium ferrite, magnetoplumbite and manganese bismuthide, all with $I_s \simeq 300$ e.m.u. or less, while a second group with high I_s consists of cobalt ($I_s \sim 1400$) and cubic-structure alloys of the Alnico type which may be caused to assume a uniaxial anisotropy when suitably prepared. There is no need to distinguish specifically between metals and the magnetic oxides, since the origin of the magnetization or the nature of the exchange energy has no direct bearing on the domain structure.

4.2. *Materials with Low Saturation Magnetization—Single Crystals*

The very simple structure of 180° domains bounded by plane parallel walls, as in Fig. 1(*a*) or 1(*b*), is in fact found only in very thin crystals; it has been extensively studied because of the possibility of testing the simple quantitative

relation developed by Kittel (2.6) or, alternatively, of measuring the wall energy. Thin crystals of barium ferrite or magnetoplumbite may be prepared with ease.

Most of the investigations have been made on the basal planes, the Bitter figure technique giving patterns like Fig. 12(*a*) (Plate III) (see Williams and Sherwood 1958) on crystals of magnetoplumbite. Extremely thin crystals are transparent (the colour ranging from pink to deep red), and similar patterns have been obtained on barium ferrite using the Faraday effect (Kooy 1957). Excellent resolution was achieved in the latter work although the components of the microscope used were widely separated and the specimens were viewed through holes in the pole-pieces of an electromagnet.

These observations showed that the energy of the simple structure is apparently less than that of any structures of parallel walls with more complex cross section. If the condition that there must be no divergence of the magnetization across the walls were rigidly maintained the same simple structure should be found in crystals of any thickness, but in fact it was found that patterns of undulating lines, as shown in Fig. 12(*b*), were formed on the basal planes of crystals about 80 μ thick (Williams and Sherwood 1958, Roberts and Bean 1954). That these lines represent the intersection of the boundaries of 180° domains running throughout the crystal is confirmed by comparing the spacing with that of the lines formed on the axial planes; Fig. 12(*c*) shows the patterns on both principal surfaces of a crystal of barium ferrite 10 μ deep.

This apparent anomaly is explained by supposing that the walls are plane in the bulk of the crystal (giving a minimum wall area for a specific spacing) but undulate with increasing amplitude throughout limited regions near the surface as shown in Fig. 12(*d*) (Goodenough 1956). The more complex pattern of free poles produced at the surface leads to a reduction in the magnetostatic energy density, and this effect more than compensates for the energy associated with those parts of the wall which no longer run parallel to the easy axis.

The undulations increase in amplitude, as expected, as the thickness of the crystals increases, but when this is about 100 μ a modified structure is formed. Patterns of isolated rings appear between the undulations as shown in Fig. 12(*e*) and these are considered to represent the intersection of small 'spikes' of reverse magnetization (Fig. 4(*b*)). Such purely surface structures also reduce the surface magnetostatic energy, without introducing as much wall energy as would be produced by cylindrical domains running throughout the whole depth of the crystal. Patterns on large crystals give similar but more complex surface structures. Fig. 12(*f*) shows some four concentric reverse spikes while the principal 180° walls are greatly distorted. The axial planes of large crystals are readily examined; the patterns represent 180° walls intersecting the surface to give straight parallel lines which on magnetization move together in pairs as described below (Bates, Craik, Griffiths and Isaac 1960).

The relation between domain spacing and crystal depth has been studied by measurements of the patterns on the basal planes of a number of crystals of magnetoplumbite (Kaczér and Gemperle 1960). The results, shown in Fig. 13, indicate that $d \propto L^{1/2}$ up to $L = 10\ \mu$ (in which range the patterns are straight lines), in accord with Eqn (2.6); the value of the wall energy given by the intercept in Fig. 13 is

$4{\cdot}82 \pm 0{\cdot}07$ erg cm^{-2}. This is very close to the value of 5·0 erg cm^{-2} found independently by measuring the spacing of parallel walls on a crystal about 6 μ in depth (Isaac 1959). These results are of particular value since precise calculations of wall energies cannot be made due to uncertainties in the values of the exchange constant A. However, such measurements cannot give both the magnetostatic energy density and the wall energy, and the validity of the expression for the former ($1{\cdot}7I_s^2 d$) has to be assumed. (The μ^* correction would, in this case, make very little difference.)

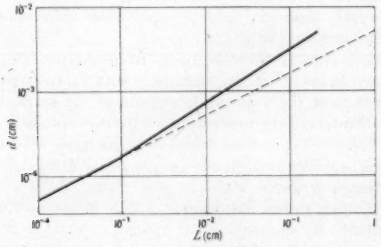


Fig. 13. Variation of the domain width d in magnetoplumbite with specimen thickness L (Kaczér and Gemperle 1960).

The point of inflection in Fig. 13 is associated with the appearance of undulating walls at the specimen surface, the undulations increasing in amplitude until $L = 50$ μ, above which the more complex patterns referred to above were found (Kaczér); in the latter range $d \propto L^{2/3}$. Since the structures vary so much there is no obvious reason why a constant relation should apply. An exact calculation of the energy and spacing must take account of the magnetostatic energy density of the complex patterns of free poles on the surface, and this is scarcely practicable at this stage. Other factors to be considered are the magnetostatic energy of the internal surfaces of the spikes and their interaction energies, and the total wall energies of the spikes.

Kaczér's measurements were confirmed independently, for the higher range of L, by measurement of the spacing of the walls on axial planes (Isaac 1959). This also confirms the view that the continuous lines in the patterns on the basal planes represent the 180° walls while the other lines represent purely surface structures.

The structures described as being typical of specimens of a particular thickness in fact only occur after either thermal demagnetization or magnetization to saturation along the easy axis. If a thin crystal of magnetoplumbite which 'normally' gives a pattern of undulating walls on its basal plane is subjected to a large field normal to its easy axis, patterns such as that shown in Fig. 14 (Plate IV) are subsequently formed. These are clearly different from those in Fig. 12(*b*) and provide a good example of the possible effects of the magnetic history on the domain structure of a specimen in the demagnetized state.

4.3. *Materials with Low Magnetization—Polycrystalline Specimens*

Each individual crystallite in a polycrystalline specimen is expected to have a domain structure basically similar to those of crystals discussed above, but with certain modifications due to its limited dimensions (where fine-grain materials are concerned) and ' interaction ' with adjacent grains.

Interaction affects the spacing in the following ways. The domain walls tend to run continuously across several grains, so as to minimize the divergence of the magnetization at the grain boundaries which are approximately normal to the easy directions (Fig. 15 (Plate II) is for a specimen of grain-oriented barium ferrite). Thus the spacing of the domains in any particular crystallite is not simply related to its own depth (as in Eqn (2.6)) but also to that of its neighbours, and a barely perceptible increase of spacing with crystal depth is found on plotting values measured with the specimen shown in Fig. 15. Significant variations in spacing are found between specimens of different average grain size, the spacing having an average value of 4 μ for an average depth of 10 μ, and about 2 μ for a depth of 2·5 μ.

Table 3. Average Diameter of Particles containing 1, 2, 3, etc., Domains (Barium Ferrite)

Number of domains	1	2	3	4	5	6	7
Average diameter (μ)	1·2	2·7	4·0	5·5	6·1	7·0	8·0

The second effect of interaction with neighbouring grains is a reduction in the effective value of the magnetization at the grain boundaries. If the latter were perfectly oriented and strain-free the divergence of the magnetization would approach zero, but in specimens of grain-oriented barium ferrite comparison of the measured spacing with that found in isolated crystals of the same thickness indicates that I_s is effectively reduced by a factor of one half.

Specimens of barium ferrite with grains 10 μ or more in diameter are relatively ' soft '. There is particular interest in specimens with smaller grains, since these have very much higher coercivity, about 2000 Oe, and remanence almost equal to saturation, and are in common use as commercial permanent magnets. The permanent magnet properties have been explained in terms of the supposed single-domain nature of small particles. Clearly no domain structure can exist at remanence in this case, but it has been found that in the demagnetized condition most of the grains of high-coercivity specimens contain two or more domains. From patterns such as Fig. 16(*a*) and 16(*b*) (Plate IV) on the axial and basal planes respectively, it is possible to compile a table showing the average diameters of single-domain and two-domain particles, etc. Since the spacing remains virtually constant the number of domains is simply dependent on the size of the particles. This is one of the few cases in which the critical size for single-domain particles can be directly demonstrated.

The contortion of the walls, as seen at the basal planes in Fig. 16(*b*), is not due to Goodenough's suggestion mentioned earlier, but to the limited dimensions of the crystallites: crystals of similar depths but unlimited lateral extent would contain plane walls.

The value of the critical particle size for manganese bismuthide is considerably greater than that for barium ferrite, no structures being found in crystals up to 30 μ in diameter (Sur 1959). The difference in the observed values corresponds to the difference in the values of the saturation magnetization of the two materials.

4.4. *Crystals with High Saturation Magnetization*

Although this section is devoted almost entirely to cobalt, the existence of similar patterns, for example, on specimens of Alcomax III, which have acquired a uniaxial anisotropy by cooling from above the Curie point in a magnetic field, indicates that the structures to be described have no special relation to the close-packed hexagonal structure.

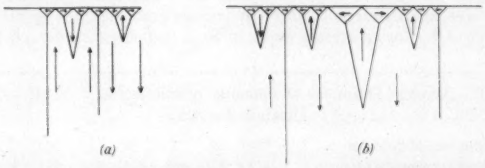


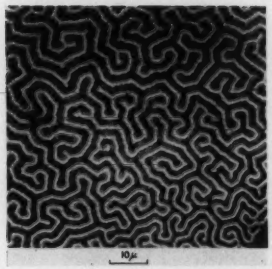
Fig. 17. Domain closure structures on a crystal of cobalt.

The simplest alternative to the ' free pole ' structure found in crystals with low saturation magnetization is the closure structure shown in Fig. 1(*d*), which must involve considerable anisotropy energy when it occurs in uniaxial crystals. More complex closure structures, such as that shown in Fig. 17(*a*) (Lifshitz 1944), may form since the introduction of the reverse spikes, which do not in this case intersect the surface, has the effect of reducing the volume of the closure domains. This reduction may be very considerable for the more complex variations, as in Fig. 17(*b*).

Patterns on the basal planes of cobalt crystals are superficially similar to those on the large crystals of magnetoplumbite, and the similarity of behaviour in small applied fields (cf. Figs 18(*a*) and 18(*b*) (Plate V)) has given the impression that they also represent free pole configurations. At higher magnifications, however, important differences are apparent, the lines of the pattern on cobalt (Fig. 19 (Plate V)) forming a continuous network which cannot be divided into long continuous walls and discrete rings.

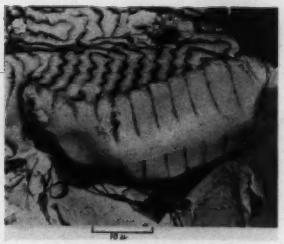
Differences are also found in the behaviour of the patterns in applied fields. On crystals with low I_s alternate strips or rings become covered by, or denuded of, colloid particles according to whether the emergent flux is parallel or anti-parallel to an applied field. In contrast to this the details of the patterns on the basal planes of cobalt crystals remain unmodified throughout a magnetization cycle, indicating that there is no normal component of the magnetization across the surfaces within the loops. It is only in fields of several thousand oersteds that the strands of the pattern become progressively thicker until patterns such as Fig. 20 (Plate V) are formed. This behaviour indicates the presence of a closure structure,

PLATE III

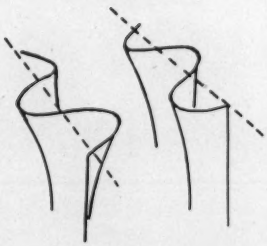


(*a*)

(*b*)

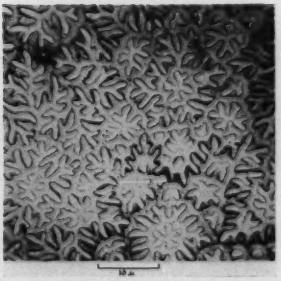


(*c*)



(*d*)

(*e*)



(*f*)

Fig. 12. Domain patterns in uniaxial single crystals with low I_s ((*c*) and (*e*) barium ferrite, the remainder magnetoplumbite). The thickness of the crystals giving the patterns in (*a*), (*b*), (*e*) and (*f*) increases progressively from 5 μ to 2 mm ; (*c*) shows patterns on basal and axial planes of a small crystallite ; (*d*) diagram of undulation in a domain wall.

Fig. 14. Domain pattern on the basal plane of a magnetoplumbite crystal, after application of a high field normal to the easy direction (× 2000).

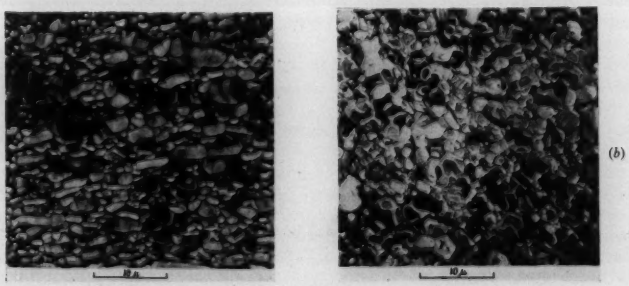


Fig. 16. Domain patterns on small grain barium ferrite (*a*) on axial plane, (*b*) on basal plane.

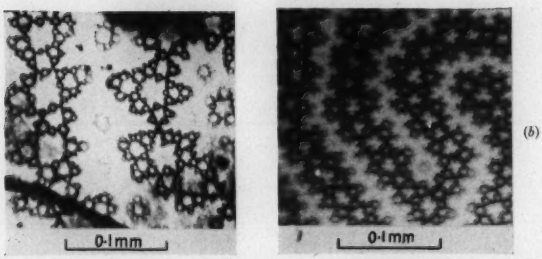


Fig. 18. Domain patterns on the basal plane of (*a*) cobalt (×200), (*b*) barium ferrite (×200), by optical microscopy.

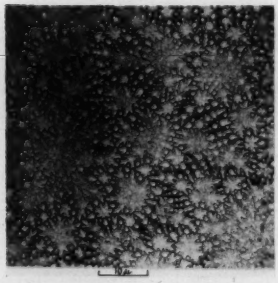


Fig. 19. Domain patterns on the basal plane of a cobalt crystal at high magnification.

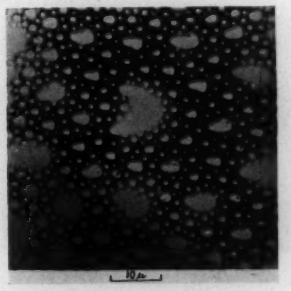


Fig. 20. Domain patterns on the basal plane of a cobalt crystal in a high field.

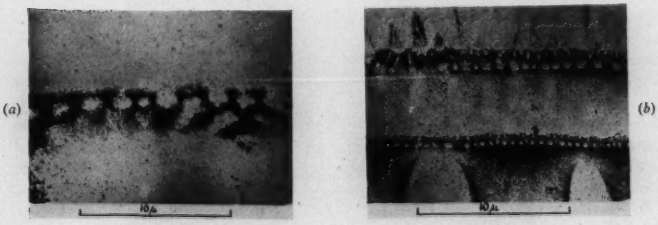
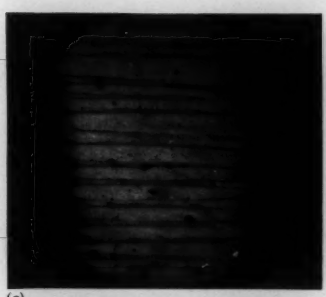
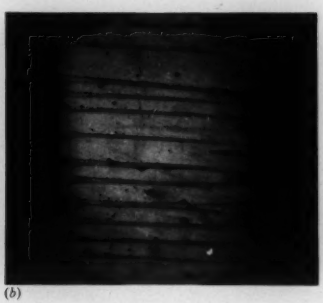


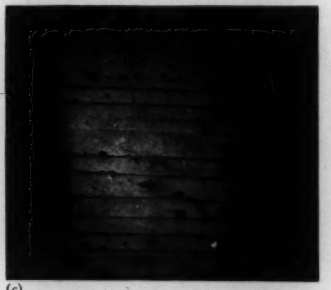
Fig. 21. Strain patterns on the axial plane of a cobalt crystal, (*a*) scratch, (*b*) slip lines.



(*a*)



(*b*)



(*c*)

Fig. 22. Formation of 180° domains and spikes on the axial plane of a magneto-plumbite crystal (Bates, Craik, Griffiths and Isaac 1960).

Fig. 23. Domain pattern on the face of a thin crystal of barium ferrite 3·0 μ thick, obtained by the Faraday effect. $H_{\perp} = 2820$ Oe (Kooy and Enz 1960).

which is only transformed to a free pole structure by the effects of high fields. It seems that this structure must be of the Lifshitz type, but in any large crystal it will be very much more complicated than those shown diagrammatically, and it would be difficult to estimate the energy associated with it.

Due to the greater experimental difficulties thin crystals of cobalt have not been as extensively investigated as those of barium ferrite or magnetoplumbite, but some striking results have been obtained in the study of cobalt whiskers, about 50 μ thick, grown by reducing cobalt bromide in hydrogen (Kaczér, Gemperle and Hauptman 1959). It was suggested that a new type of structure existed in these crystals. The domain spacing was measured where the convergence of two faces gave a varying specimen thickness and it was found that the spacing varied as the two-thirds power of the depth. The usual Lifshitz closure structure was still formed in crystals of the order of 10 μ deep, although a simple free pole configuration is indicated for depths of 5 μ or less. It seems that a half-power law applies in a lower thickness range down to about 0·7 μ below which the spacing reaches a minimum value and then increases.

Fine-scale patterns in the form of chains with loops 1 μ or less in diameter (Fig. 21(*a*) (Plate V)) are frequently found on the axial planes of cobalt crystals. These have been shown to be caused by scratches which have been largely removed by electro-polishing, and are considered to depend on residual stresses. Similar patterns also form in a more ordered manner as in Fig. 21(*b*) along slip lines on crystals which have been plastically deformed, and are similarly interpreted with respect to the stresses persisting at the slip-planes. No such patterns are found on barium ferrite or magnetoplumbite, and this difference is probably due to the difference in the mechanical hardness rather than to any difference in magnetic properties.

4.5. *The Magnetization of Uniaxial Crystals*

When a field is applied parallel to an easy direction the 180° domain walls move to permit the growth of alternate domains in which the magnetization is directed parallel to the field, at the expense of the less favoured domains. On axial planes of cobalt, pictures have been obtained showing that the walls remain parallel throughout their movement (e.g. Elmore 1938 b), but more recent work has indicated that this does not represent the behaviour of the internal domain structure.

It has already been noted above that the patterns on the basal planes of cobalt crystals persisted virtually unchanged throughout the hysteresis cycle, and similar results have been obtained for crystals of low I_s. Magnetization measurements commonly deal only with the flux through the central regions of the specimens, and the study of the basal plane patterns illustrates the approximate nature of the term ' technical saturation '. The presence of these patterns provides an explanation of the fact that relatively low fields can produce such ' saturation ' since the surface structures greatly reduce the demagnetizing fields which would otherwise need to be overcome. It is also clear that nucleation is not necessarily involved in the magnetization cycle.

The behaviour of the 180° domains in the bulk of the crystals must also be governed by the structures at the basal planes, since the walls are effectively fixed

at their extremities by the intervening structures near the basal plane. When the patterns on accurately aligned axial planes of cobalt crystals are studied, the walls remain straight and parallel during their movement, but their spacing is not that of the internal domains, since they meet and disappear when the crystals are magnetized to about only one-third of saturation. Although the reason is not apparent, the behaviour of walls on axial planes of magnetoplumbite specimens (Bates, Craik, Griffiths and Isaac 1960) corresponds more to the expected behaviour of internal domains. Here it is seen that the equally-spaced parallel lines which correspond to the demagnetized state become curved when a field is applied, so that they approach in the central regions but maintain the same spacing at the ends (Fig. 22(*a*) (Plate VI)). As bulk magnetization proceeds, the pairs of lines become sensibly parallel, about 1 μ apart, over most of their length (Fig. 22(*b*)), after which they abruptly transform into pairs of spikes (Fig. 22(*c*)), which in turn diminish in size and eventually disappear.

Thus it appears that magnetization is opposed, not by the increase in demagnetization energy produced by a uniform pole distribution over the end faces, but by demagnetizing fields over the internal domain surfaces where boundaries diverge from the easy directions, as well as by a small increase in wall energy. The process has yet to be studied quantitatively.

The magnetization of extremely thin crystals of barium ferrite, in which no superficial structures are formed, is more likely to be effected by the parallel approach of alternate pairs of walls, with consequent increase in magnetostatic energy due to unequal pole distribution. This is indicated by Fig. 23 (Plate VI), obtained by the Faraday effect with partially magnetized crystals (Kooy and Enz 1960). The process does not consist simply of changes in the width of the individual domains, but of changes in the double spacing, and the variation of this parameter agrees with calculations based on changes in the magnetostatic energy. ' Unfavourable ' domains disappear in an applied field a little below $4\pi I_s$, and after saturation new domains always appear first at one of a limited number of fixed points in the specimen.

The relatively large value of single-domain particle size in manganese bismuthide facilitates the direct demonstration of the rotation of the magnetization. This is indicated (in a crystal, diameter 20 μ) by the progressive movement of the Bitter figure deposits around the edges of the crystal, when fields up to 24 000 oersteds are applied normal to the easy direction (Sur 1959).

The behaviour of crystals which are a little above the critical size is of particular interest (Sur, Stol'c, Kandaurova and Bulatova 1957). A field of 6000 oersteds along the easy axis eliminates all structures observed on the axial plane and when the field is reduced to zero new domains appear and spread gradually across the surface. After the application of 21 000 oersteds no new domains appear as the field is reduced to zero and raised to 3500 oersteds in the opposite direction; a further very small field increment then causes the colloid which has been spread uniformly over the surface to migrate rapidly to the edges of the specimen. This was taken as evidence of a rapid rotation of the magnetization, but it is also possible that domains which nucleate in the high field spread throughout the specimen too rapidly to give any detectable patterns.

§ 5. Domain Structure in Iron

5.1. *Iron Crystals*

While uniaxial crystals almost invariably have a basic structure of 180° domains, the six easy directions in iron and eight in nickel allow for more complex structures, which are strongly dependent on the shape of the specimen and on its magnetic history, a factor generally of little account in uniaxial materials.

Figure 24(*a*) (Plate VII) shows a pattern on a (100) surface (Bozorth 1951) in which the striations indicate that the walls separate 180° domains magnetized parallel to a [100] direction. The triangular formation suggests the formation of the simple prismatic closure structure of Fig. 1(*d*) and determinations of the spacing of the domains and the corresponding crystal depths (measured along the

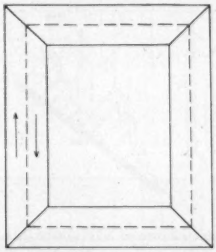


Fig. 25. ' Picture-frame ' crystal with 180° boundary.

direction of the magnetization) agree with calculation. The 180° walls are frequently decorated by ' fir-tree structures ', when the surfaces on which they are observed diverge by a few degrees from planes containing the easy axis ; these have been accounted for in terms of a partial surface flux closure, i.e. a reduction of the magnetostatic energy arising from the normal component of the magnetization across the mis-oriented surface (Fig. 24(*b*)) (Bozorth 1951).

When a field is applied parallel to the length of the domains, alternate pairs of walls approach one another, and magnetization proceeds by the growth of those domains magnetized in the same direction as the field. No quantitative studies of this apparently simple mode of magnetization appear to have been made ; the behaviour of the closure domains would clearly be of great importance and might be very difficult to analyse.

A crystal cut in the shape of a picture-frame (Fig. 25) is particularly suitable for quantitative studies, as it has no appreciable demagnetizing factor, and once saturated, it tends to remain so. When a field is applied to one of the arms a single 180° domain wall may nucleate to effect demagnetization as shown. These specimens are particularly adapted to studies of the impedance of walls by local inhomogeneities or inclusions, and changes in magnetization have been related directly to the observed movements of walls (Williams and Shockley 1949). In the absence of an applied field, the walls have a tendency to move towards the centre of the specimen so as to decrease the total wall energy.

Complex patterns with fine spacings are found on surfaces which do not contain an easy direction (e.g. (111)) and a connection between the two types of pattern is suggested by the echelon structures found near the edges of (100) surfaces and drawn in Fig. 26 for a quadrant of a thin disk (Martin 1957). A gradual transition from the simple closure structure to a complex arrangement of small 180° and 90° domains, which reduce the surface energy where the edge of the disk makes a large angle with the (100) surface, is indicated. Since the edge of the disk at the mid-point of the quadrant is a (110) surface, containing an easy direction normal to the plane of the drawing, subsidiary domains may also form along this direction, but may be considered less likely the thinner the crystal. If the surface were a (111) plane (inclined at an angle to the plane of the drawing) the magnetization could only run parallel to it at the expense of increasing anisotropy energy.

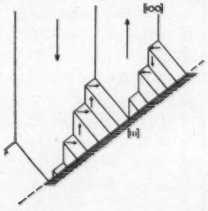


Fig. 26. Echelon structure on a (111) face (iron) (Martin 1957).

Qualitatively, echelon structures may be compared with the surface structures in uniaxial materials, in that they reduce the surface magnetostatic energy by producing a pattern of free poles, which is more finely divided than that corresponding to the internal domain spacing ; alternatively, as in cobalt, they reduce the volume which is occupied by closure domains magnetized out of the easy direction. Calculations of the spacing of the principal domains, as a function of the crystal depth parallel to the directions of magnetization, have been made by assuming that the spacing of each part of the structure is independent of that of the other parts (Martin 1957) ; the values obtained experimentally with some crystals agreed within a factor of two with those calculated.

In the above examples, 90° domains have occurred which although subsidiary may constitute a more substantial part of the general structure. The occurrence of 180° and 90° walls is related both to the form of the crystals (see below) and to the past magnetic treatment. Where substantial holes or inclusions exist, small closure domains can reduce or eliminate the corresponding magnetostatic energy (Fig. 4), and if in the course of magnetization a 180° wall moves across such an inclusion, it is found that the transverse domains can grow (Néel 1944 a). These investigations revealed several processes : on the reduction of an applied field from a saturation value, *transverse* (Néel) spikes first form at imperfections, and as these grow, the intersection of the 90° wall with inclusions produces 180° walls which in turn give rise to 90° walls (Bates and Martin 1956). The ' sub-domains ' at imperfections have also been studied by Sur and Abels (1955 a, b, c), who found

that on reduction of the saturating field these transverse sub-domains reappear before the surface domains, which take the form of reverse spikes. On further reduction of the field, extensive 180° structures develop from these reverse spikes, which are found at the surfaces of single crystals and on grain boundaries in polycrystalline specimens; no 180° structures develop from the sub-domains at imperfections (Sur and Abels 1955 d, see also Goodenough 1954).

In general, a demagnetized iron crystal may be said to be in a six-phase state, the magnetization in the various domains being distributed between all six easy directions, and magnetization proceeds by a reduction in the number of easy directions occupied. For example, Fig. 27 (see Williams 1952) demonstrates how the volumes of one set of 'unfavourable' 90° domains may be reduced by the movement of a set of 180° walls. The most comprehensive investigations of

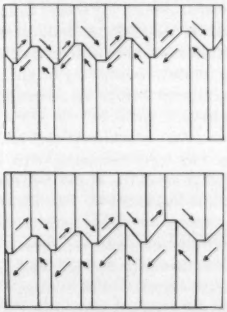


Fig. 27. Movement of a 180° boundary, with increase in volume of one set of 90° domains and decrease in the volume of the adjacent domains (Williams 1952).

domain processes and magnetization have been concerned with crystals of the form shown in Fig. 28(*a*) (Plate VII), with a structure, principally of 90° domains, which was postulated by Néel (1944 b, c), who derived the corresponding theoretical magnetization curves; earlier theories of the magnetization curve had not been related to specific domain structures (Akulov 1929, Lawton and Stewart 1948).

The existence of structures corresponding to those of Fig. 28(*a*) has been verified by the Bitter figure method (Bates and Neale 1950), but while the variation of the spacing of the main domains was approximately as predicted, the value of the spacing was greater by a factor of up to three. Further investigations (Bates and Mee 1952) revealed that at the (110) face the domains designated q in Fig. 28(*a*) (Plate VII) were of the simple form indicated, but the p domains were further sub-divided to give a complex free-pole pattern at the surface: the 'lace patterns' of Fig. 28(*b*).

Lee (1953) modified Akulov's theoretical magnetization curve (1929) for the [110] direction, taking into account the small closure structures as well as the principal domains, and obtained a set of magnetization curves with the actual

volumes of the p and q closure domains as parameter. In a more advanced investigation (Bates and Hart 1956 a) experimental magnetization curves were compared with those calculated by Lee, and these were further corrected to correspond to the observed volumes of the closure domains. The correlation was considerably improved although mainly over a restricted field range. This approach, which depends on theoretical relations for a postulated model being subsequently corrected to conform more closely to the observed structures, seems to be the only one possible, and appears to be justified by the results obtained.

5.2. *Whiskers*

The reasons for the general complexity of the patterns on iron crystals have been indicated and it is doubtful whether the complete domain structure of any sizeable specimen apart from the ' picture-frame ' can be deduced. Simple patterns which give very clear demonstrations of certain aspects of domain structures and magnetization have been obtained on the minute iron crystals which are formed as ' whiskers ' by, for example, the reduction of ferrous chloride in hydrogen. These have the great practical advantages of possessing bright, flat and strain-free surfaces, which require no special treatment before examination and are generally parallel to simple crystallographic planes.

The patterns obtained on (100) whiskers (Scott and Coleman 1957) show that they generally contain only two 180° domains with simple closure at the ends (Fig. 29(*a*) (Plate VIII)), but if a field is applied perpendicularly to the long axis and then removed, the structure indicated by Fig. 29(*b*) (and Fig. 30 (Plate VIII)) is found. Both of these structures are stable with no divergence of the magnetization, the latter structure (29(*b*)) being stable because the original 180° structure has been destroyed. Any structure such as Fig. 29(*c*), for example, would certainly not be stable in a homogeneous crystal. The volume of the transverse domains as shown in Fig. 29(*b*) can only be reduced when the 180° walls are caused to move in a longitudinal field (Fig. 29(*d*)). The bulk of these crystals becomes saturated in moderate fields (*c.* 200 Oe), but Scott and Coleman have shown by the Bitter figure method that closure structures persist up to about 600 Oe at the ends.

The demagnetizing factors of these long narrow crystals would be very low if they were considered to approximate to ellipsoids of appropriate eccentricity, but in fact the local demagnetizing fields are very different and may be very high at the sharp corners (Shtrikman and Treves 1960 a). A very careful examination, by the Kerr magneto-optic technique modified to give high resolution, of the regions near the tips of whiskers with (100) faces showed that very fine domains of reverse magnetization persisted in fields up to about 500 Oe ; even with fields up to 2000 Oe transverse spikes were found, and the unusual structure shown in Fig. 29(*e*) remained up to at least 5000 Oe (Fowler, Fryer and Treves 1960). As the field was reduced from 5000 Oe, transverse spikes reappeared and at lower fields reverse domains formed, to effect eventual demagnetization.

The structure shown in Fig. 29(*e*) is not a closure structure, but exists to reduce the demagnetizing fields which would otherwise appear at the sharp corners, and demonstrates the non-uniformity of these fields. The presence of this structure may well have wider implications in the study of large crystals of iron ; there is no essential difference between the tip of a whisker and the presence of a hole or

inclusion of irregular shape in a large crystal, and since it is not customary to apply fields of many thousands of oersteds in the course of the hysteresis cycle, minute transverse domains may not be eliminated. Thus a nucleation process need not be postulated to account for the presence of transverse domains.

The domain structures of whiskers with hexagonal section and elongated in the [111] direction have also been studied (Coleman and Scott 1958), the structures in the central portions consisting of six 90° domains divided by walls which intersect at the axis, and a quantitative explanation of the closure structures found at the ends of these whiskers was given by Kaczér and Gemperle (1959). Zigzag, or serrated walls running along the sides of (100) whiskers, elastically bent (De Blois and Graham 1958), have been interpreted in terms of a model of 90° walls, by Kaczér and Gemperle (1959).

5.3. *Polycrystalline Specimens*

Polycrystalline specimens in which a crystallite surface may have any orientation are not expected to give patterns which can be readily interpreted and related to underlying structures, and a great variety of purely surface structures is to be expected. These may be extreme modifications of the ' fir-tree ' patterns, or be similar to the (111) patterns, and are not likely to give any direct indication of the structure at internal grain boundaries.

Simple and regular patterns are obtained on the ' grain-oriented ' iron strips, produced by cold rolling, which are of wide commercial application. In these the crystallites generally have an easy direction parallel or near to the rolling direction, and some of the surfaces have a (110) orientation, i.e. they have a (110) [100] texture. It has been found that the patterns on many of the grains are very similar to some patterns on single crystals (Bates and Hart 1953), and by studying the reflections from etch pits the patterns have been grouped according to the angular deviation θ of the [100] direction from its projection in the surface (Paxton and Nilan 1955). When $\theta \sim 2°$, parallel lines representing 180° domains are formed. The spacing varies with the crystallite depth (measured parallel to the easy direction) and alternate pairs of walls approach each other on application of a field along the easy, or rolling, direction. As in barium ferrite the walls frequently run continuously from grain to grain.

When θ lies between 2° and 4° patterns of spikes are formed, and on magnetization these decrease in length while maintaining, in the earlier stages, a fixed spacing at their bases. The direction in which the spikes point is related to the surface orientation as for uniaxial crystals, i.e. they point ' downhill ', and it has been inferred that similar spikes exist within the material, but pointing in the opposite direction.

At greater values of θ (4°–7°) ' lozenges ' are formed ; rows of discrete structures elongated in the [100] direction and arranged in rows perpendicular thereto. These are similar to certain patterns found on single crystals (Bates and Hart 1953) and have been interpreted as representing the intersection with the surface of rows of transverse spikes, originating at inclusions during demagnetization (Bates and Davis 1956). As they resemble the isolated elliptical figures found on some slightly mis-oriented surfaces of uniaxial crystals, it seems possible that they might represent the intersection of reverse spikes originating at internal grain boundaries.

There are many indications that the basic structure is similar to that of uniaxial polycrystals, consisting of 180° domains or reverse spikes of triangular cross section based on grain boundaries. This is not the structure expected, or demonstrated, in single crystals of iron, and the difference corresponds to that between free surfaces and grain boundaries. At a free surface some sort of closure is almost inevitable since there would otherwise be a very high pole density even after taking into account the μ^* effect: at a grain boundary normal to an easy direction there may be very little divergence of the magnetization and thus a simple 180° structure, involving the appearance of free poles of low density, may have the lowest energy. No patterns indicating the existence of closure domains are seen at grain boundaries, and the continuity of the structures across the boundaries indicates that the flux is more or less continuous rather than closed within each grain.

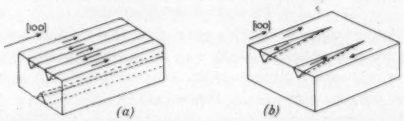
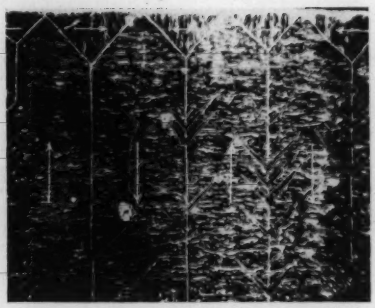


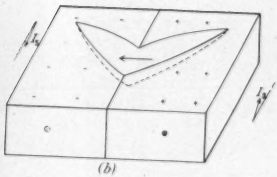
Fig. 31. (*a*) 180° domain structure at a (110) surface; (*b*) tapered domains produced at surface making a small angle with the (110) plane and [100] direction.

Later work has reinforced this interpretation of the spike pattern (Paxton and Nilan 1955). The energy of a 180° wall in iron varies according to the orientation of the plane it occupies (whilst remaining parallel to an easy direction), being less when this is a (100) than when it is a (110) plane (Table 1). Taking the second anisotropy constant into account, calculations have indicated that 180° walls in crystals with (100) surfaces should occupy planes normal to the surface, but when the surface is (110) the walls should make angles of $\pm 32°$ with a plane normal to the surface; these considerations have been verified by noting the positions of the intersections of selected walls with the surfaces of thin specimens (Graham and Neurath 1957, Harada and Tatsumoto 1959). This suggests a wall pattern such as Fig. 31(*a*) rather than one of domains of rectangular section. The structure reveals tapered outlines when the surfaces are mis-oriented (Fig. 31(*b*)), just as a pattern of spikes replaces the parallel walls on uniaxial crystals.

In randomly oriented specimens closure is more likely at some of the grain boundaries, where there may be little direct interaction between the patterns in adjacent grains; spacings measured on individual crystallites which give simple patterns agree with an expression derived for an echelon structure at internal grain boundaries, i.e. $d = 1{\cdot}32\,(\gamma/K)^{1/4} L^{3/4}$ cm (Bloor and Martin 1959). Measurements on 'cube-textured' silicon–iron (Bates and Carey 1960 a, Carey 1960), in which the crystallites are substantially aligned so that there is no separate structure at grain boundaries, indicate that the exponent in the expression above may be close to unity, i.e. $d \propto L$. The uncertainty is not reduced by the observation that if the effective value of I_s is reduced by a factor 1/5, to allow for continuity of flux across the grain boundary, the results can be brought into agreement with Eqn (2.6), i.e. $d \propto L^{1/3}$ (see also Spacek 1956, 1959 a, b).

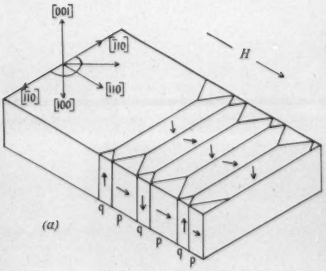


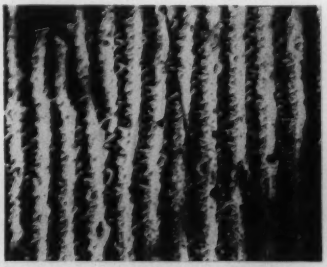
(*a*)



(*b*)

Fig. 24. (*a*) Pattern on a (100) surface of an iron crystal (Bozorth 1951). (*b*) Formation of 'tree' pattern ; the crystal surface makes a small angle with the easy direction of magnetization (I_s) and free pole is produced at the surface ($+++, ---$). The magnetization in the 'tree' domain lies parallel to the surface so that no free pole is produced.





(*b*)

Fig. 28. (*a*) Domain structure of a Néel block (Néel 1944 c) ; (*b*) lace patterns on p and q domains (P. F. Davis, private communication).

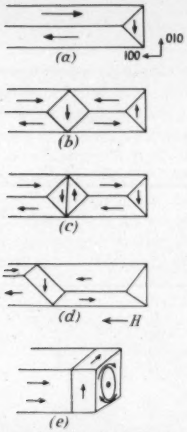


Fig. 29. Domain structure of iron whiskers (after Scott and Coleman 1957).

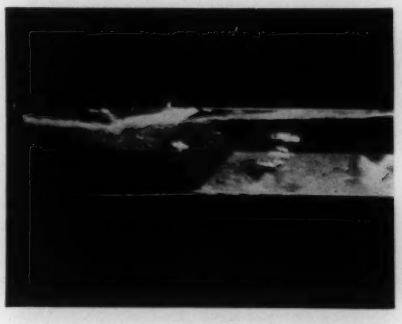


Fig. 30. Domain pattern on an iron whisker 140 μ across, corresponding to Fig. 29(*b*) (Kerr effect) (Fowler, Fryer and Treves 1960).

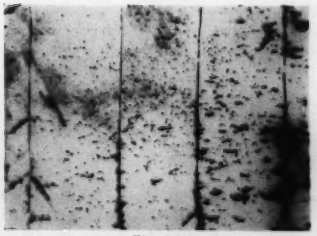


Fig. 34.

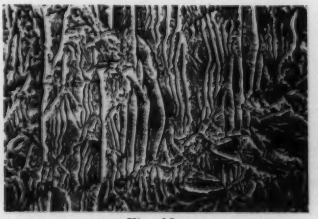


Fig. 35.

Fig. 34. 180° domains on a (110) surface of nickel (×100) (Yamamoto and Iwata 1953).

Fig. 35. Domain pattern on a surface not containing an easy axis (×112) (Yamamoto and Iwata 1953).

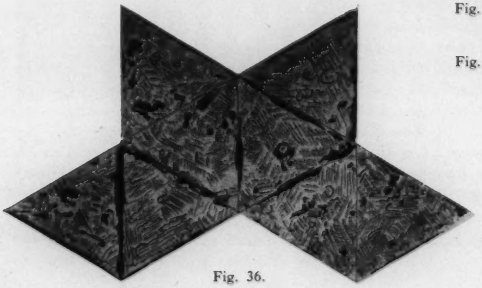
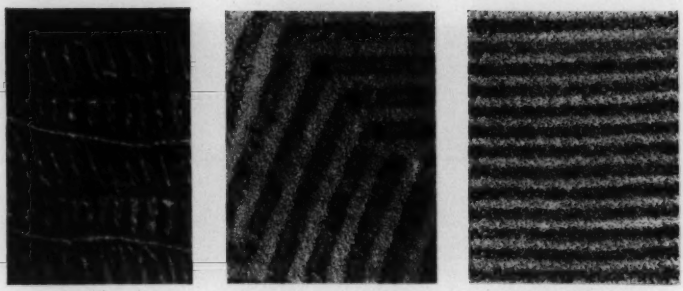


Fig. 36.

Fig. 36. Tracing of patterns on all faces of an octahedral crystal of magnetite Sugiura 1951).



(*a*) (*b*)

Fig. 37. (*a*) Patterns on (100) plane of a manganese ferrite crystal, (*b*) higher magnification (Bates, Craik and Griffiths 1958).

Fig. 38. 180° wall and 'tree structure' on (110) of manganese ferrite (Craik and Griffiths 1959).

(*a*) (*b*)

Fig. 39. Patterns on (*a*) (100) and (*b*) (110) surface of a crystal of $Mn_{0.5}Zn_{0.5}Fe_2O_3$ (Craik and Griffiths 1959).

(*a*) (*b*)

Fig. 40. Patterns on polycrystalline nickel ferrite grain size (*a*) 12 μ, (*b*) 2·5 μ with field normal to surface.

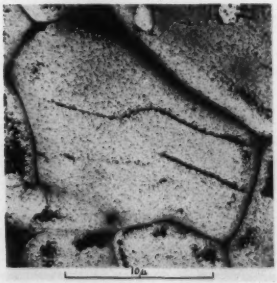


Fig. 41. Pattern on a simple grain of polycrystalline nickel ferrite.

(*a*) (*b*) (*c*) (*d*)

Fig. 43. Movement of a boundary across a single grain of polycrystalline manganese–zinc ferrite (Knowles 1960).

The domain structure in very fine polycrystalline iron does not appear to have been studied; but it has been shown that when the thickness of polycrystalline strips is reduced to about 5 μ, no structure is visible even in crystallites some 25 μ in diameter (Abels and Sur 1956), which are presumably single domain particles.

No extensive account of the important topic of wall-impedance can be given here, but reference may be made to recent observations on the behaviour of tie-domains on 'tubes' when a domain wall moves across an inclusion (Bates and Carey 1960 b).

§ 6. Domain Structure of Nickel-like Materials

6.1. *Nickel and Face-centred Cubic Alloys (Single Crystals)*

Since nickel has a face-centred cubic structure, with eight easy directions parallel to the cube diagonals, very different patterns to those on iron are expected on corresponding crystal surfaces. Differences in detail are also to be expected in the patterns on surfaces containing easy directions (the (110) surfaces of nickel replacing the (100) surfaces of iron). The diagram in Fig. 32(*a*) shows the type

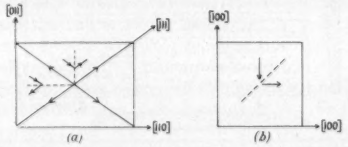


Fig. 32. Domain walls (broken lines) in (*a*) nickel and (*b*) iron crystals.

of wall which conforms to the easy axes of nickel, the magnetization turning through angles of 71° and 109° (and 180°), while the corresponding walls in iron crystals are given in Fig. 32(*b*).

In the examination of nickel some experimental difficulty is occasioned by its greater sensitivity to stress. The effect of stress depends, *inter alia*, on the relation between the magnetostriction and crystal anisotropy constants. If the magneto-elastic energy is comparable with the crystalline anisotropy energy, i.e. if $\frac{3}{2}\lambda_s\sigma \sim K$, the domain structure will be greatly affected, and while the magnetostriction of nickel is higher than that of iron its crystal anisotropy is lower by a factor of five. Thus, while surface strains due to mechanical polishing can be readily overcome, it is necessary to take great care to avoid extensive deformation.

The low value of the crystal anisotropy causes the walls to be broad and to attract colloid particles rather weakly; indeed, the first demonstration of domain structures in a face-centred cubic crystal relied on the higher anisotropy (-2×10^5 erg cm^{-3}) of 60% cobalt–nickel (Bozorth and Walker 1950). Patterns on (110) surfaces showed domains in which the magnetization vectors made all the predicted angles. Clear patterns on nickel itself were demonstrated in a number of independent but almost simultaneous investigations (Yamamoto and Iwata 1950, Martius, Gow and Chalmers 1951, Williams and Walker 1951, Bates and Wilson 1951), followed somewhat later by others (Yamamoto and Iwata 1953, Stephan 1955, Iwata and Yamamoto 1956).

The magnetization in nickel follows the easy directions, but it is more difficult to relate the patterns to internal domain structure than it is with iron, for the patterns are very dependent on both the orientation of the surfaces and the way in which the crystals are demagnetized.

The patterns on (110) surfaces indicate the predominance of 180° or 70° domains, while those on (100) surfaces of iron indicate mainly 180° or 90° domains. In both cases, tree patterns form when the surface is slightly mis-oriented ; although the small superficial domains in nickel serve the same purpose as those in iron they are generally more elongated and, as the deviation of the surface increases they do not become broader but form secondary branches with different directions of magnetization. Complex formations which reduce surface magnetostatic energy very effectively are favoured by the low wall energy.

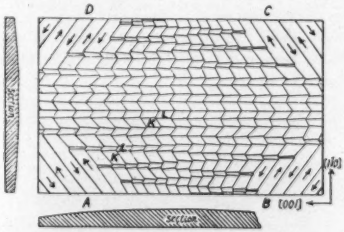


Fig. 33. Formation of parallelogram—net patterns by 180° and 109° walls : the central portion coincides with a (110) plane (Yamamoto and Iwata 1953).

Patterns comprising both 180° and 71° walls may also form on (110) surfaces, and the dependence of the predominant type of structure on the precise surface orientation has been demonstrated on a surface which was accurately oriented in the central region but deviated from the (110) plane towards the corners due to the effects of electro-polishing (Yamamoto and Iwata 1953). In the middle all four easy directions may be effective without involving any divergence of the magnetization across a surface, but towards the corners only the directions approximately normal to the line of maximum slope can be used, and here 180° domains are present (Fig. 33). On these grounds alone, however, there is no reason why one of the sets of 180° domains should not continue across the middle, but the four sets of 180° domains at the corners exert a magnetostrictive stress on the central portion of the crystal which is most easily accommodated by the formation of a slightly distorted parallelogram net structure.

However, the parallelogram net pattern is only found after demagnetization by an alternating field applied in a direction midway between two of the easy axes ; when the field has been applied parallel to an easy axis the corresponding set of 180° domains, as shown in Fig. 34 (Plate VIII), run right across the crystal, in spite of the divergence in I_s incurred at two of the corners. It would be difficult to determine which type of structure represents the lower energy, as different terms are involved. It is also uncertain whether some of the structures are confined to regions near the surface or run throughout the depth of the crystal.

Since (211) surfaces contain only one easy axis, 180° domains alone can form without considerable divergence of the magnetization across the surface, and the gradual transition from parallelogram net to 180° patterns has been illustrated during an extensive examination of general surfaces on a hemispherical crystal (Iwata and Yamamoto 1956). Surfaces which do not contain an easy axis, such as those parallel to (100) or (111) planes, give patterns of dense branching lines with spacing equal to that of the simpler patterns on (110) surfaces, between which fine closure structures are closely packed (Fig. 35, Plate VIII).

Smaller crystals or crystals cut to special shape might give structures which could be studied quantitatively, but such specimens might be difficult to prepare in a strain-free condition. However, the particular effects of the negative crystal anisotropy appear to have been satisfactorily demonstrated, together with the importance of magnetostrictive stresses.

6.2. *Domain Structure of Cubic Ferrites—Single Crystals*

This section is concerned chiefly with ferrimagnetic materials, with the general formula $MOFe_2O_3$, which have a spinel–cubic crystal structure and negative crystal anisotropy (with easy directions in the [111] directions). As far as domain structure is concerned the nature of the anisotropy is more important than the chemical composition or general physical properties (justifying the consideration of materials such as barium ferrite in a different section).

Ferrites generally have lower values of saturation magnetization than metals, ~300 e.m.u., and, if mixed ferrites of the formula $M_xA_{1-x}Fe_2O_3$ are included, the crystal anisotropy ranges from -10^3 to -10^5 ergs cm^{-3} (Bozorth, Tilden and Williams 1955). Cobalt ferrite is exceptional in having positive (cubic, iron-like) crystal anisotropy, and the addition of small amounts of cobalt to other ferrites may greatly affect the anisotropy, extending the range to $+10^6$ ergs cm^{-3}, but little attention has been paid to the ferrites with positive anisotropy.

Despite intensive investigations of the general properties of ferrites, to some extent connected with their wide commercial applications, their domain structures seem to have been inadequately studied. As with the ferromagnetic metals these studies have their inherent interest as they may be correlated with the magnetic constants of the materials ; a knowledge of the domain structure of representative ferrites might also be useful in connection with work on magnetothermal effects (e.g. Bates and Christoffel 1957), high speed reversal of magnetization (Menyuk and Goodenough 1955) and ferrimagnetic resonance (e.g. Smit and Wijn 1954) in which quantitative predictions are often based on the spacing and behaviour of postulated domain structures. The majority of such investigations concern polycrystalline or sintered specimens with a range of grain size, and it is by no means certain that coherent domain structures always exist in such materials. The paucity of the results on ferrites may be ascribed to early difficulties in the preparation of strain-free surfaces on materials which cannot be electro-polished. The following account, therefore, contains some experimental detail.

In the first demonstration of domain structure in ferrites, these difficulties were avoided because certain crystals of magnetite (Fe_3O_4 or $FeO.Fe_2O_3$) have flat and lustrous surfaces which require no special preparation (Sugiura 1951). These crystals have the form of regular octahedra, whose (111) surfaces do not

contain any easy direction, but from the study of the patterns on all eight surfaces it has been inferred that the internal structure consists of walls parallel to (110) planes (Fig. 36, Plate VIII). The similarity of the patterns to those on corresponding surfaces of nickel crystals (cf. Fig. 35) indicates the existence of the same type of domain structure.

Crystals of magnetite and of nickel ferrite shaped into hollow parallelograms with arms parallel to [111] directions have been shown, after careful mechanical polishing and etching, to have simple domain structures with walls running parallel to the edges (Galt 1952, Galt, Andrus and Hopper 1953).

When a single crystal boule of manganese ferrite, grown by the flame fusion method, is cleaved as a preliminary to surface orientation, two distinct types of pattern are found (Bates, Craik and Griffiths 1958). The surface approximates to a (100) plane, and the large-scale patterns shown in Fig. 37(*a*) (Plate IX), indicating an internal spacing of about 50 μ in a crystal 1 cm in diameter, have been interpreted in a similar way to those on (100) surfaces of nickel. At higher magnification there appears a pattern of very fine light and dark bands or simply lines of similar spacing which is less than the predicted width of a wall (Fig. 37(*b*)). The large-scale pattern consisted of periodic intensification of the fine structure and the latter is thought to represent a superficial structure arising from very high stresses induced in the surface layers during cleaving.

Surfaces oriented to (100) and (110) planes have been obtained by grinding and polishing, and luckily they etch in hot concentrated acids to give bright flat surfaces. Domain structures analogous to those on nickel are then seen, 180°, 109° and 71° walls being found on the (110) surface, with associated tree structures when this is slightly mis-oriented (Fig. 38 (Plate IX)) (Craik and Griffiths 1959). Patterns have also been obtained on a polished and etched (110) surface and the two adjacent natural (111) surfaces of a small octahedral crystal of nickel ferrite (grown in a borax melt). When a field is applied in a [110] direction on a (110) surface, 109° domains lying normal to the field direction are seen, with closure structures on the (111) surfaces having the same spacing and direction, suggesting an overall structure comparable with that of the Néel block of Fig. 27(*b*).

In general the observations confirm the expected similarity of the domain structures to those of nickel, although in one case at least there are important differences in detail. The patterns on a manganese ferrite crystal are often confused by the departure from linearity of the intersections of the walls with the surface. They are frequently curved over a considerable length, or sharply kinked; presumably the energy associated with the divergence of the magnetization across these walls is compensated by a reduction of wall energy due to local inhomogeneities in the crystal. The walls will tend to run selectively through regions in which either the exchange energy or the crystalline anisotropy energy is lowered by variations in chemical composition or crystallographic inhomogeneity.

These irregularities in the structure are likely to be more pronounced in materials with lower anisotropy (or saturation magnetization) and studies of manganese–zinc ferrites should be of value. Only one crystal of $Mn_{0{\cdot}5}Zn_{0{\cdot}5}Fe_2O_3$ (expected to have $K_1 \simeq 3 \times 10^3$ ergs cm^{-3}) seems to have been investigated. It gives patterns like those in Fig. 39(*a*) (Plate IX) on a cleft (100) surface, and in Fig. 39(*b*) on a polished and etched (110) surface. Ignoring

the superficial fine structure in the former pattern, these indicate a uniaxial structure, and it is assumed that stresses set up during the formation or cooling of the crystal must have favoured one pair of easy directions to the exclusion of the others; stresses of the order of 10^7 dyn mm^{-2}, giving $\frac{3}{2}\lambda_s \sigma \sim K$, would be required.

The examination of strain-free crystals with very low anisotropy (below 10^3 ergs cm^{-3}) may well be impractical by the Bitter figure method, since the very wide walls will have little tendency to attract colloid. Nevertheless, the great wall width is one of the most interesting aspects which call for investigation, and it is fortunate that the methods using polarized light are not so limited.

6.3. *Polycrystalline Cubic Ferrites*

Polycrystalline or sintered ferrites present special problems of surface preparation.

It is easy to observe characteristic domain structures in large-grain barium ferrites after careful mechanical polishing (§ 4.3), although this treatment obscures the structures on very fine grains; because of the low anisotropy of the cubic ferrites such polishing tends to give strain patterns which are not characteristic of the internal structure (Bates, Clow, Craik and Griffiths 1958). The strain patterns can be removed by etching, but this is not satisfactory, since grain boundaries are selectively etched away and the generally porous material is attacked internally, leaving surfaces pitted.

A method of surface preparation which seems to be generally satisfactory was suggested by the manner in which ferrite compacts are produced. After careful mechanical polishing with diamond compounds to give a flat and lustrous surface the specimen is raised to a suitable sintering temperature for a few minutes. Representative patterns on high-anisotropy ferrites are found to be much more sharply defined after such treatment than after mechanical polishing alone; the surface stresses could be relieved by sintering for a time sufficiently short to avoid detectable grain growth or change in magnetic properties. A fortuitous advantage of this method is that the grain boundaries are shown in slight relief and can be easily observed in the optical microscope or by the replica method in electron micrographs.

This and similar methods have been used in the examination of barium, nickel and cobalt ferrites with medium or fine grains (Craik 1959); a similar method was developed independently for the specific study of manganese–zinc ferrite (Knowles 1960). In the latter case, the surface preparation is associated with the preparation of the specimen itself: pre-sintered powders are pressed by punches with specially polished faces which impart such a flat surface to the finally sintered specimen that no further preparation is required. This appears to be the more direct method when the preparation of the specimen is under control.

Figures 40(*a*) and 40(*b*) (Plate X) show patterns on specimens of nickel ferrite which have average grain sizes of 12 μ and 2·5 μ, respectively. There are three types of pattern on the first specimen. Many of the grains are covered by complex patterns with a spacing of less than 0·5 μ, while simple line patterns with spacing of 2–3 μ appear on surfaces of similar extent. The difference in spacing alone suggests that the patterns of the first kind represent closure or superficial free-pole structures on surfaces such as (100) or (111), while the simpler patterns form on

surfaces containing one or more easy axes and are thus approximately representative of the domain structures of the whole of the grains and, in a general way, that of the internal grains. The third type of pattern, dark patches with rather diffuse boundaries, is assumed to represent the intersection of domains in which the magnetization makes a small angle with the surface. (A small field was applied normal to the surface.)

The lines of the simple patterns are rarely straight, and may be sharply kinked as illustrated by Fig. 41 (Plate X). As in manganese–zinc ferrite such irregularities are considered to correspond to regions of inhomogeneity, where wall energy is locally reduced; this point is referred to further in studies of manganese–zinc ferrite below. It is doubtful whether quantitative analysis of such patterns would be justified, since even if the directions of the magnetization were known the spacing would be strongly dependent on the structures in neighbouring grains.

As the average grain size decreases, somewhat simpler structures are found, and on the fine-grained specimen the patterns indicate two-domain structures in grains 3 μ in diameter, while no sub-division occurs when the diameter falls below 2 μ, the ' critical diameter ' for the compacted state, as distinct from that which may apply to isolated particles.

Wall movements cannot be followed directly by replica electron microscopy, but limited optical observations on the coarse-grained specimen indicate that structure frequently changes discontinuously, and occasionally the structure on a particular grain undergoes a sudden complete rearrangement when fields are applied. A much fuller investigation of domain behaviour in a grain of a manganese–zinc specimen with grains some 50 μ in diameter has been undertaken using optical microscopy (Knowles 1960). The specimen was in the shape of a ring and had a square loop, with squareness ratio about 70%.

The grain particularly studied had a rectangular section at the surface and was divided by two principal walls into what were assumed, from their behaviour and from the geometry of smaller structures, to be 180° domains running parallel to the field direction. Making other reasonable assumptions, the magnetization of the grain as known fields were applied was derived from the areas of its surface intersected by the favoured domains. Taking the field acting on the grain to be the general internal field, a hysteresis loop was then drawn and compared with that of the specimen as a whole, as in Fig. 42. The important changes in the magnetization, giving the distinctly square shape, are illustrated by Figs 43(*a*) to 43(*d*) (Plate X) corresponding to the points 8, 10, 11 and 12 on the loop. In spite of the relatively large field change from 8 to 10, little wall movement occurs. A small further increase in the field then causes one section of the wall (the upper part in Fig. 43) to break free from an apparently inhomogeneous region, while the other part remains stationary in a region in which it is considerably distorted (Fig. 43(*c*)) and a further small field increment frees this latter portion (Fig. 43(*d*)).

A general treatment based on these observations gives loops for the whole specimen which correspond well with those measured; although many assumptions are necessary, there seems to be little doubt that this kind of wall impedance is of fundamental importance.

There does seem to be considerable scope for further investigation of the domain structure and magnetization processes in polycrystalline ferrites. According to the

above considerations, specimens with truly homogeneous and strain-free grains should have very different characteristics from that of the specimen discussed here, and factors which are of importance in one particular specimen may not be important in others with greatly different anisotropy.

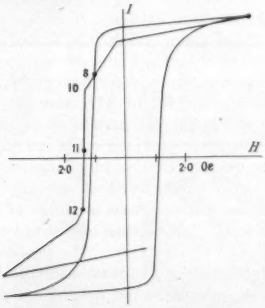


Fig. 42. Hysteresis loop for single grain of polycrystalline manganese–zinc ferrite (cf. Fig. 43) (Knowles 1960).

6.4. *Rare Earth Garnets and Ortho-ferrites*

The ferrimagnetic garnets have the general formula $3M_2O_3 5Fe_2O_3$ and the ortho-ferrites $MFeO_3$ where M is a rare earth or yttrium ion. Single crystals of these materials can be produced which are stoichiometrically nearly perfect and they are thus preferable to the ferrites for general studies on ferrimagnetism.

Table 4. Crystallographic and Magnetic Data for Garnets of the Type $5Fe_2O_3.3M_2O_3$

The saturation magnetization is given in Bohr magnetons per unit formula $5Fe_2O_3.3M_2O_3$. The compensation temperature is that where the magnetization due to one set of ions is balanced by the equal and opposite magnetization due to the others, giving a net magnetization zero (Bertaut and Pauthenet 1957).

Ion M	Y	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
Compensation temperature θ_c (°K)	—	290	246	220	136	84	$4 < \theta_c < 20$	—	—
Curie temperature θ_f (°K)	560	564	568	563	567	556	549	548	549
Saturation magnetization (μ_b)	9·44	30·3	31·4	32·5	27·5	23·1	2·0	0	8·30

The crystal structure of the garnets is cubic, of a rather involved form (space group $O_h(10)$–Ia3d, see Geller and Gilleo 1957) with 160 atoms per unit cell. The crystals have faces which are either (211) or (110). The magnetic properties of the garnets are summarized in Table 4, taken from Bertaut and Pauthenet (1957). The magnetic anisotropy K is low, of the order 10^4 erg cm^{-3}. The yttrium garnet

(YIG) and gadolinium garnet (GdIG) have been the subject of intensive investigation because of their application in microwave devices. At room temperature the saturation magnetization I_s of YIG is about 140 e.m.u.; because of the presence of a compensation point near room temperature the value of I_s for GdIG is only about 3 e.m.u.

Since garnets are transparent to light at a thickness of the order of 50 μ, they can be investigated by the Faraday effect. In the work of Dillon (1958 a, b), selected crystals were ground down to a suitable thickness, 100 μ, with (112) or (110) surfaces. Despite the care taken in the final polishing the patterns obtained, e.g. those on the left of Fig. 44(*a*) (Plate XI), were ascribed to residual stresses. Crystals in which the magnetization ran parallel to the surface, as on the right in Fig. 44, showed light and dark bands which could not be due to Faraday rotation, and were considered to be due to magnetic birefringence.

Domain patterns can also be obtained on crystals using the normal Bitter pattern method. In YIG for instance these are often of a very high standard and rather more distinct and clearly defined than one would expect from the low value of the magnetic anisotropy ($2K/I_s \sim 10^2$ Oe). The possibility arises that these may be due to residual surface strain since it appears that annealing and surface polishing can produce very noticeable changes in the spacing and form of the patterns.

The magnetization of the ortho-ferrites $MFeO_3$ is generally very weak, and they have a crystal structure of the perovskite type. From early domain structures it was suggested that the magnetization followed the (111) directions but later work, supported by torque measurements, has indicated that the ortho-ferrites are uniaxial with the orthorhombic *c* axis magnetically favoured; the magnetic anisotropy is high (see below).

In an extensive investigation (by Sherwood, Remeika and Williams 1959) on yttrium and eleven rare earth ortho-ferrites, the Bitter figure method was used for large crystals and the Faraday effect for thinner crystals. Unpolished crystals were studied and there was no evidence of stress anisotropy; the results indicate that the magnetization lies along the orthorhombic *c* axis and gives rise to simple 180° structures of large spacing. Fig. 45 (Plate XI) shows a surface normal to this axis, on which the intersections of the walls are represented by the broad lines which divide the crystal into three domains only (a three-domain crystal of barium ferrite would be less than one hundredth of this size), and Fig. 46 (Plate XI) shows a similar structure observed by the Faraday effect. The significant features of these structures are their extreme simplicity and great spacing, as compared, for example, with barium ferrite. The domain spacing in a barium ferrite crystal of corresponding depth would be smaller by a factor of about ten, and such simple structures would only be formed in crystals about 1 μ thick; these differences can be qualitatively explained by the low value of the magnetization. Applying the Kittel formula (Eqn (2.10)) for simple 180° domains the wall energy may be roughly estimated from Fig. 47 (Plate XII) as approximately 3 erg cm^{-2} which is what might be expected for a ferrimagnetic crystal with high anisotropy.

In these investigations on the ortho-ferrites, it was found that wall movement could be observed in applied fields of less than one oersted while saturation was achieved in fields of 50 Oe. Once saturated the crystals generally remained so until reverse fields of about 2500 Oe were applied, whereupon the magnetization

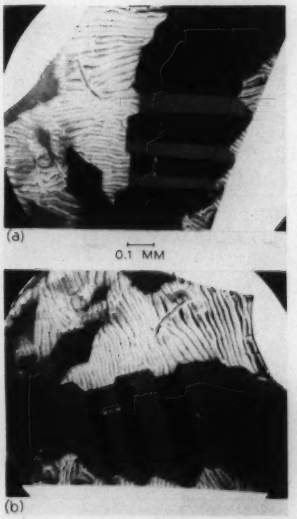


Fig. 44. Domain patterns in IG crystals 100 μ thick (Dillon 1958 b).

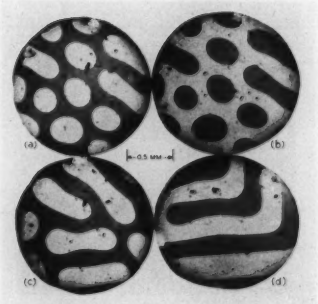


Fig. 46. Domain structure on the surface normal to the orthorhombic axis in an yttrium ortho-ferrite crystal (Sherwood, Remeika and Williams 1959).

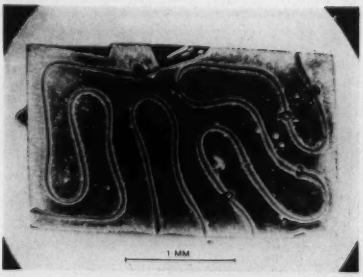


Fig. 45. Domain structure on the surface normal to the orthorhombic axis in an yttrium ortho-ferrite crystal (Sherwood, Remeika and Williams 1959).

Fig. 47. Domain structure on surfaces normal and parallel to the orthorhombic axis of ortho-ferrite crystal (Sherwood, Remeika and Williams 1959).

Fig. 49. Domain pattern on the surface normal to direction of magnetic annealing field — Alcomax III (×1500) (Craik and Isaac 1960 a).

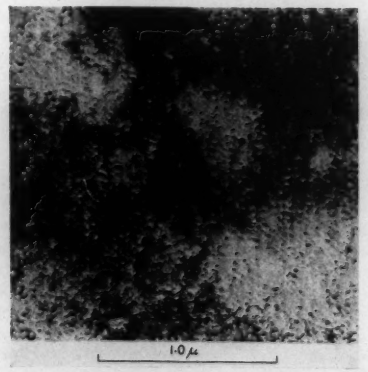


Fig. 50. Fine structure on a surface normal to the direction of the annealing field on Alcomax III in its optimum condition (×35 000).

PLATE XIII

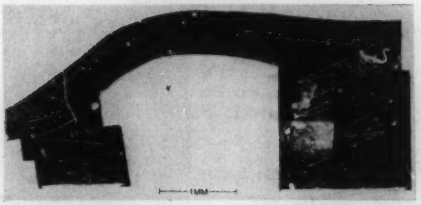


Fig. 52. Composite photograph showing a domain pattern on a 300 Å film of Fe–Ni–Mo. The base of the picture is the edge of the film (Williams and Sherwood 1957).

Fig. 53. Domain pattern showing nuclei of reversal magnetization on the edge of a 1300 Å film of Fe–Ni–Mo. Easy axis vertical (Williams and Sherwood 1957).

Fig. 54. Domain pattern on a 2000 Å film of iron. Easy axis vertical (Williams and Sherwood 1957).

Fig. 62. Domain pattern showing double boundaries on a 147 Å Fe–Ni–Mo film after demagnetizing with field 90° to direction of easy magnetization. Easy axis approximately vertical (Williams and Sherwood 1957).

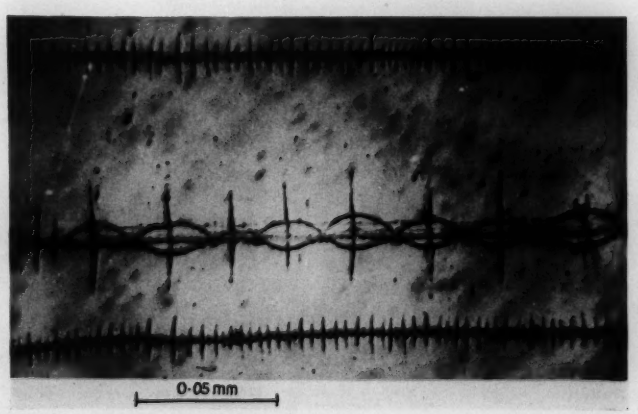


Fig. 58. Cross-tie wall and chain wall in Permalloy film of 500 Å thickness (Gomi and Odani 1960).

was abruptly reversed, and the crystals could not generally be demagnetized by a cyclic field. This behaviour (as with that of the fine-grained barium ferrite) is typical of a material for which the nucleation field is very much higher than the field required for saturation. That the anisotropy of the ortho-ferrites was indeed high, giving very high values of K/I_s, was shown by further experiments in which fields up to 18 000 Oe caused no appreciable rotation of the magnetization away from the c axis. The coercive fields might be expected to be even higher than the observed values if the crystals were wholly strain-free and regular in shape. It was found that if the crystals were strained by mechanical polishing, high reverse fields were not required and in some instances multi-domain structures formed spontaneously on reducing the applied field to zero.

§7. Alloys which Respond to Magnetic Annealing, and Powders

When certain binary or tertiary alloys of iron, nickel and cobalt are cooled from above their Curie point in an applied field they develop a uniaxial anisotropy in the field direction. Some alloys also acquire, on cold rolling, uniaxial anisotropy which should be apparent in their domain structures. In the quenched state these alloys may have such a low anisotropy that coherent domain structures, as opposed to closed flux loops, may not be formed. Reference is first made to investigations of a series of nickel–cobalt alloys annealed without a magnetic field.

7.1. *Annealed Nickel–Cobalt*

The numerical values of the anisotropy of these alloys decrease with increasing nickel content as shown in the second line of Table 5. The polycrystalline specimens examined are very sensitive to strain, but after careful polishing and

Table 5. Value of the Anisotropy Coefficient K for Nickel–Cobalt Alloys (from Yamamoto and Taniguchi 1956)

H_1, magnetic field limit for observed boundary movement; H_2, magnetic field limit from magnetothermal measurements.

% Cobalt	21	29	46	59
$K \times 10^{-6}$ (erg cm^{-3})	−2	−30	−70	−130
H_1 (Oe)	—	30	75	140
H_2 (Oe)	30	50	120	200

annealing, domain patterns indicative of negative anisotropy have been obtained, and the changes produced by measured effective fields have been studied (Bates and Isaac 1959). Wall movement was observed, on all but the first specimen, up to the field limits shown in the third line of Table 5. The fourth line gives the field limits below which it appears, from a consideration of magnetothermal measurements, that magnetization occurs by wall movement (Bates and Clow 1959). There is reasonable agreement for all but the first specimen, but it is not possible to say whether the failure to observe domain structures directly was due to their absence or to the failure of the Bitter figure technique.

7.2. *Perminvar*

The effect of allowing ring-shaped specimens of Perminvar (43% nickel, 34% iron, 23% cobalt) to cool slowly from above the Curie point in fields of about 15 Oe applied around the ring are well illustrated by the hysteresis loops in Fig. 48. The magnetic anneal raises the initial and maximum permeability (to 400 000 or more), and induces a square loop. Specimens which have been cooled slowly without applied field have the peculiar waisted *minor* hysteresis loops in which dB/dH changes rapidly near the origin.

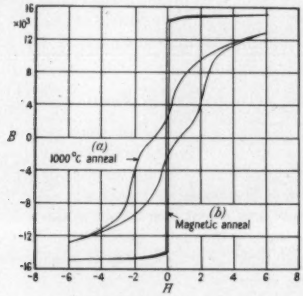


Fig. 48. Hysteresis loops of a Perminvar ring, after (*a*) 1000°C anneal, (*b*) subsequent magnetic anneal (Williams and Goertz 1952).

The domain structure of such rings in the magnetically annealed state is unusual in that the walls are curved, running continuously round the rings (Williams and Goertz 1952). There need not, however, be any divergence of the magnetization across the walls if the magnetization itself around the ring follows closed loops. The patterns of Williams and Goertz show clearly that the walls are curved within the grains, and not polygonal with changes in direction at the grain boundaries.

The square hysteresis loop (of Fig. 48) can be explained in terms of the observed domain structures, and the shape of the specimens. When a saturating field is reduced to zero there is no tendency towards demagnetization. A finite reverse field is required to nucleate reverse domains, whose walls sweep across the specimens to reverse the magnetization completely, since there are no demagnetizing fields to be overcome.

The square loop might also be explained in terms of rotation of domain vectors, but this is eliminated by the observation (Williams and Goertz 1952) that if the field at which the magnetization begins to reverse is rapidly reduced the change is halted, and a ' re-entrant ' hysteresis loop is obtained. The domain walls then remain, and their impedance to free movement is further demonstrated by the rounded corners of minor hysteresis loops. If it were not for a finite wall impedance no domain structure would be stable, as the walls would tend to collapse towards the inner rim of the specimen. A further peculiarity of such specimens is that their magnetization curves should depend upon the site of nucleation : Williams

and Goertz found that cutting a notch on the outside rim of the specimen decreased the coercive force and hysteresis loss by a factor of two.

The domain structures of annealed specimens of Perminvar itself do not appear to have been studied. It may be assumed that the direction of magnetization varies from grain to grain, so that a square loop is not to be expected, but the reasons for the waisted loop are by no means apparent and are discussed in the next section.

7.3. *Cobalt–Nickel Alloys*

The domain structures of crystal of 40% cobalt–nickel alloy, which exhibits all the characteristics of Perminvar but is easier to examine owing to its somewhat higher anisotropy, have been examined by Yamamoto, Taniguchi and Aoyagi (1956, 1958) following a theory of magnetic annealing put forward independently by Taniguchi and Yamamoto (1954) and Néel (1954). According to this theory the anisotropy induced by magnetic annealing is due to an anisotropic orientation of the spins approximately along the direction of the magnetization. This is, of course, the direction of the applied field if it is sufficient for saturation. Thus, magnetic annealing should only be appreciable when the Curie point is high enough for there to be a certain mobility of the atoms in the temperature range immediately below it. Even in the absence of a field, the uniaxial anisotropy should be induced by slow cooling at any point within such a crystal, but the directions will then correspond to those of the domain structure and may well give an overall cubic anisotropy.

Rod-shaped crystals, elongated in an easy (111) direction, were studied after various heat treatments. It was found that a crystal which had been annealed and cooled slowly had a complex structure of 71°, 109° and 180° domain walls, while crystals which had been either quenched rapidly from above the Curie point or cooled slowly in a field had more widely spaced structures of 180° walls parallel to the long axis. It was inferred that the fine structure of the slowly cooled specimen was that which had existed just below the Curie point, at which stage the mobility was greatest and the anisotropy was induced within each domain ; when the specimen was cooled further there could be no rearrangement to what would otherwise have been a structure of lower energy, since this would have involved rotation against the induced anisotropy. No anisotropy was induced in the quenched specimen, and the structure of 180° walls parallel to the long axis was that of lowest energy for the shape of specimen. It appeared to be coincidental that the crystal cooled in the field had a similar structure : a system of 180° walls running diagonally across the axis of the rod would presumably have been set up if the field had been applied parallel to a different easy direction.

From this work it seems that complex structures in which all the easy directions are occupied, also exist in pure metal crystals just below the Curie point, but in these cases rearrangements may occur freely at lower temperatures. The ' freezing in ' of the complex structures probably occurs quite generally in cubic solid solutions. The cobalt–nickel alloy referred to earlier (§ 6.1) (Bozorth and Walker 1950) as well as the annealed nickel–cobalt alloys mentioned in this section (Bates and Isaac 1959) would be subject to this effect. ' Characteristic ' domain structures of solid solutions, in fact, only exist in quenched specimens.

It seems that all the important properties associated with Perminvar can be explained in terms of these three distinct types of domain structure. Quenched

specimens have hysteresis loops of ' normal ' shape, with no waist, and the individual structures arrived at in the demagnetized state, although of the same type, may vary considerably in detail. In the specimens which have been cooled slowly, however, there is only one structure which represents a very low energy. In the initial stages of magnetization the walls must be moved from positions of low energy, and thus the initial permeability is low compared with that of the quenched specimen. Again, when the magnetization approaches zero, assuming that full saturation has not been too closely approached, the walls are attracted to their positions of low energy and dB/dH thus increases to give the very low remanence and coercivity, and the waisted or constricted loop. Finally, when these specimens are saturated, the hysteresis loops are not constricted since the destruction of the fine-scale structure eliminates the possibility of the system of walls simultaneously approaching their positions of minimum energy. Renewed annealing is necessary before even the minor loops resume their constricted shape.

7.4. *Precipitation Hardened Alloys*

The alloys of which Alcomax III† may be taken as representative (' Alnico ', ' Ticonal ', etc.) also acquire uniaxial anisotropy when allowed to cool in a magnetic field. They may however acquire a much higher coercivity than alloys of the Perminvar type, particularly when subjected to further heat treatment (the ' optimum ' state). The properties of these and other permanent magnet materials have recently been reviewed (Wohlfarth 1959).

Polycrystalline specimens which have been field-cooled only have domain structures of arrays of 180° walls, the directions of which generally make small angles with the cooling field direction. (This divergence of the directions of the domains may only arise because the applied field is insufficient to cause true saturation.) Patterns such as that in Fig. 49 (Plate XII) (Craik and Isaac 1960 a), which are obtained on surfaces normal to the cooling field direction, are very similar to the patterns on the basal planes of cobalt crystals, and are quite distinct from those on barium ferrite. They thus indicate the formation of closure domains of the Lifshitz type (1944) which correspond to the high value of the saturation magnetization. If any very fine precipitate exists in specimens in this condition it has no directly detectable effect on the domain structure.

The high coercivity (about 600 Oe) in specimens of Alcomax, in which precipitation has developed to a considerable extent, has been explained in terms of the fields required to cause rotation of the magnetization in the individual precipitate particles (Stoner and Wohlfarth 1948). The rotation, in the single-domain particles, is opposed by the crystal anisotropy or the shape anisotropy of the elongated particles. This does not involve any domain formation, and appears to be incompatible with the clearly-defined domain structures which have in fact been observed in specimens with high coercivity (Nesbitt and Williams 1950, Andrä 1956 c).

The domain structure of Alcomax in its ' optimum ' condition differs in certain details from that of the material which has been field-cooled only. The patterns on surfaces parallel to the cooling-field direction generally consist of uniform sheets of colloid with no intensification at the edges, such as generally indicates the presence

† Al 8%; Ni 13·5%; Co 24–25%; Cu 3%; Nb 0·3–1%.

of the domain walls, and the domains are generally much narrower. The patterns on surfaces normal to the field direction indicate a free pole structure, and the colloid deposits are not uniform but have a fine internal structure, shown in Fig. 50 (Plate XII), which is not present in patterns on the corresponding surfaces of barium ferrite or of cobalt. It has been suggested, from such observations, that the domains observed on this material are not of the usual type in which the individual magnetic carriers are maintained in alignment by exchange forces; but are regions in which the individual magnetic moments of precipitate particles are mutually aligned by magnetostatic interactions (Craik and Isaac 1960 a). Exchange forces would still be operative within the particles, and would form the basis of the ferromagnetism. However, it is difficult to speak with confidence of Alcomax itself since both the matrix and the precipitate are probably ferromagnetic, and it may not be justifiable to consider the precipitate particles as isolated in considering exchange forces.

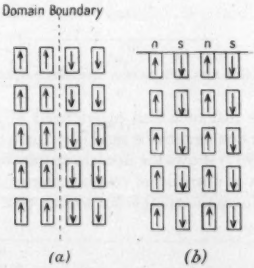


Fig. 51. Alternate domain formation in single domain particle material.

7.5. *Powder or Elongated Single Domain Magnets*

It is interesting to find that domain structures which behave on magnetization in the manner expected for uniaxial materials can be demonstrated on powder, or elongated single domain magnets (Craik and Isaac 1960 b). These consist of very fine, elongated particles of iron or an iron–cobalt alloy, which are coated with tin and embedded, with their long axes aligned, in a non-magnetic matrix (Luborsky, Paine and Mendelsohn 1959). The purpose of the coating is specifically to prevent ' exchange ' forces from acting between particles, and on this basis the structures observed have been interpreted as ' magnetic interaction domains '.

There should be important differences between the boundaries of this special kind of domain and the conventional domain wall, in which both exchange and anisotropy energy are involved (§ 2). The boundary between two interaction domains (in the perfect array represented by Fig. 51(*a*)) is simply a plane between two adjacent layers of particles and its energy depends only on the regularity of the array of the particles and the consequent effectiveness of the flux-closure between them; in a specimen in which the particles were perfectly aligned the boundary energy might be so low that the structure with least total energy would approach that shown in Fig. 51(*b*).

§8. Thin Films

The domain structure of thin films is of some importance, not only because of their application as switching elements, but because of the relative uniformity with which they can be made and the correspondingly simple form of the associated magnetostatic energy terms.

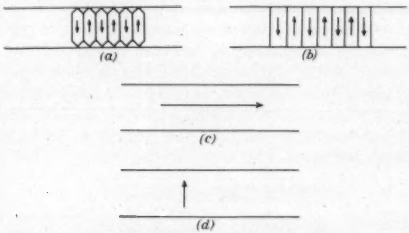


Fig. 55. Domain structures in a thin film with an easy direction perpendicular to the plane of the film.

An obvious example is that of a film magnetized in a direction parallel to the plane of the film so that the demagnetizing field is virtually zero. Fig. 52 (Plate XIII) (Williams and Sherwood 1957) shows the domain structure in such a film and Fig. 53 (Plate XIII) the nucleation of domains of reverse magnetization ; the formation of a zigzag boundary which reduces the energy due to free-pole on the boundary is shown

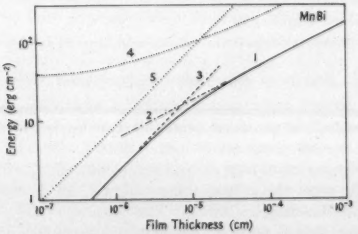


Fig. 56. Variation of domain energy of thin film with thickness (Málek and Kamberský 1958).

in Fig. 54 (Plate XIV). The change in magnetization is achieved by (*a*) a sideways movement of the long 180° boundary, (*b*) movement of the zigzag in the direction of magnetization and (*c*) expansion of a domain of reverse magnetization.

The domain structure of ferromagnetic films is modified as the thickness is decreased because of the relatively high energy associated with the formation of the domain boundary. In general, below a critical thickness this energy is so high that the domain structure disappears, but considerable variations in the process are possible.

The simplest structures are those illustrated in Fig. 55(*a*)–(*d*), where the film has an easy direction of magnetization at right angles to the plane of the film. For thicker films the structures (*a*) or (*b*) are to be expected for materials of low or high magnetic anisotropy respectively. Kittel (1946) obtained expressions for the domain energy of structures (*a*), (*b*) and (*c*) and showed that below a thickness of the order of 3×10^{-5} cm the domain structure takes the form of a uniformly magnetized film parallel to the surface (*c*).

Table 6. Energy Density of Bloch Walls (*a*) and Néel Walls (*b*) for Massive Material and Thin Films for the Limiting Case of Zero Thickness of Film $\gamma_0 = \sqrt{(AK)}$ (Kaczér 1957)

Type of wall	Crystallographic orientation	Scheme	Energy density of wall (ergs cm^{-2}) Massive material: Generally	Massive material: For iron	Thin film: Generally	Thin film: For iron
180° (*a*)	(100)		$2\gamma_0$	1·6	$2\sqrt{\left(\frac{8\pi I_s^2}{K}\right)}\gamma_0$	21
180° (*b*)	(100)		$2\sqrt{\left(\frac{8\pi I_s^2}{K}\right)}\gamma_0$	21	$2\gamma_0$	1·6
90° (*a*)	(100)		γ_0	0·8		
90° (*a*)	(110)		$1{\cdot}73\gamma_0$	1·4	$\sqrt{\left(\frac{8\pi I_s^2}{K}\right)}\gamma_0$	10·5
90° (*b*)	(110)		$0{\cdot}304\sqrt{\left(\frac{8\pi I_s^2}{K}\right)}\gamma_0$	3·2	γ_0	0·8
180° (*a*)	(110)		$2{\cdot}76\gamma_0$	2·2	$2\sqrt{\left(\frac{8\pi I_s^2}{K}\right)}\gamma_0$	21
180° (*b*)	(110)		$2\sqrt{\left(\frac{8\pi I_s^2}{K}\right)}\gamma_0$	21	$2{\cdot}76\gamma_0$	2·2

For materials with high uniaxial anisotropy the formation of closure domains may require a relatively high energy contribution so that structure (*b*) may be preferred. The curves in Fig. 56 (Málek and Kamberský 1958) show the variation with thickness of the energy of the domain structures (*b*), (*c*) and (*d*) for MnBi ($K \sim 8{\cdot}5 \times 10^6$ erg cm^{-3}); with a film thickness of less than 100 Å the width of the domain is so great that the specimen is virtually uniformly magnetized.

The problem is complicated by the fact that there is also the possibility that it may be energetically preferable for the magnetization vector in a boundary wall to rotate as shown in Fig. 57(*b*), i.e. in a plane at right angles to the wall (cf. the ' Bloch ' wall in Fig. 57(*a*)) so that ' free-pole ' effects (div $I \neq 0$) are produced in

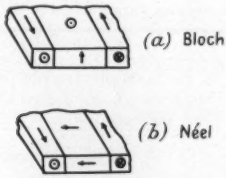


Fig. 57. Representation of the rotation of the magnetization vector in (*a*) a 180° Bloch wall with the production of free pole on the surface of the film and (*b*) a Néel wall with ' free pole ' in the wall (div $I \neq 0$) (Kaczér 1957).

the boundary (a Néel wall). Kaczér (1957) has examined this problem and the results are collected in Table 6. For iron it is seen that for thin films the energy of a Néel wall is much less than if the wall structure were associated with ' massive ' material.

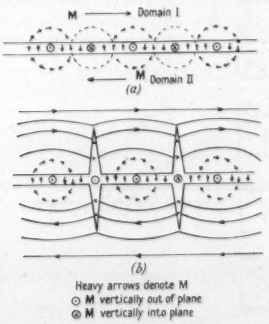
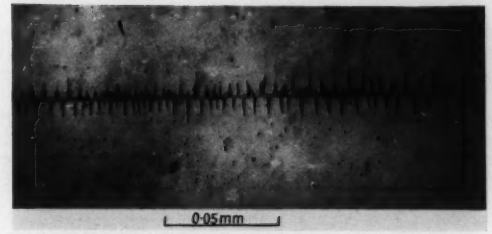
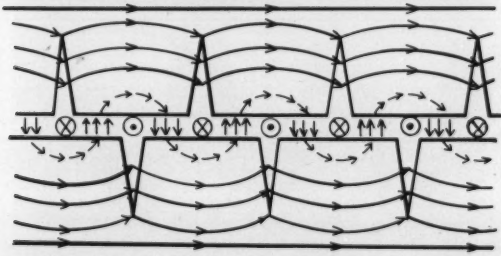


Fig. 59. Flux distribution at a 180° domain wall in the plane at a low-anisotropy film (*a*) without, (*b*) with, the formation of ' cross-tie ' walls (Huber, Smith and Goodenough 1958).

The formation of ' cross-tie ' walls has been observed in thin films. Huber, Smith and Goodenough (1958) have shown that the flux closure necessary to reduce the magnetostatic energy can only be brought about at alternate intervals along the boundary as in Fig. 58 (Plate XIV), and that this structure allows for continuity in the flow of induction (Fig. 59). Displaced cross-tie walls have also been observed (Fig. 60 (Plate XV)) and these have been shown to be caused by the

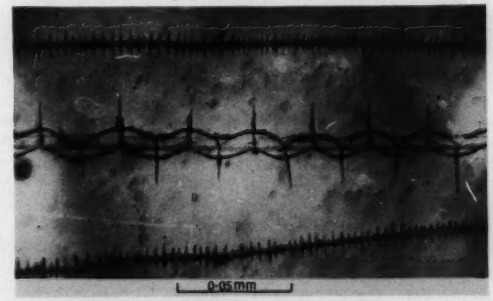


(*a*)

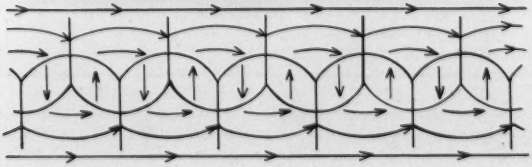


(*b*)

Fig. 60. (*a*) Displaced cross-tie wall (circled points, ⊙ vertically out of plane; circled crosses, ⊗ vertically into plane) and (*b*) its magnetic structure (Gomi and Odani 1960).



(*a*)



(*b*)

Fig. 61. (*a*) Displaced chain wall and (*b*) its magnetic structure (Gomi and Odani 1960).

magnetization vectors on each side of the boundary being parallel (and not anti-parallel as is usual). Under certain conditions (in regions of greater strain) ' chain ' walls may also be produced (Figs 58 (Plate XIV) and 61 (Plate XV)) (Gomi and Odani 1960).

When the thickness of the film is greatly reduced double walls may be observed (Fig. 62, Plate XIV) and these have been discussed by Kaczér (1958) on the basis of the dipole interaction between the free pole induced on the surface by neighbouring walls.

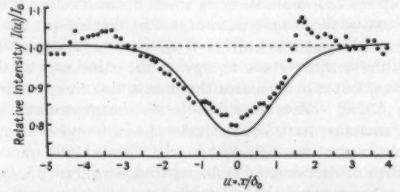


Fig. 63. Measured (000) and calculated intensity distribution of image of domain wall produced by electron transmission. x = distance from centre of boundary (Fuller and Hale 1960).

The thickness of domain walls in thin films has been investigated using the electron microscope technique of Fuller and Hale, described in § 3.3. The domain wall in a 200 Å, 50–50 Permalloy film was photographed and the photograph scanned using a micro-densitometer ; the resultant intensity distribution for the dark and the light wall images is shown in Fig. 63. The wall width is given as approximately 1500 Å, which compares favourably with estimates of the width.

§ 9. Domain Walls

Relatively few direct measurements of the width or energy of domain walls have been reported. Reference has already been made to the measurement of the 180° wall energy in magnetoplumbite (Kaczér and Gemperle 1960) which agreed closely with the calculated value, and also to the direct measurement of wall-widths in thin films by transmission electron microscopy (Fuller and Hale 1960). An indication of the width of walls in large crystals may be obtained from electron micrographs of powder patterns : for example, while the edges of patterns obtained in normally applied fields on cobalt ferrite or barium ferrite are extremely sharply defined, with a resolution or ' spread ' of about 10^{-6} cm, the edges of patterns from cobalt crystals are always more diffuse. Since the anisotropy constants of these materials are similar a significant difference in their exchange constants is indicated and an independent measurement of the wall energy in cobalt would thus be of considerable value. A method of measuring wall-width from the change in intensity of a light beam which is caused by the Kerr effect (Kirenski and Veter 1959) may be of potential value for the relatively broad walls which are to be expected in materials with low anisotropy. The results given for 180° walls in silicon–iron, about 10^{-4} cm, are greater than the calculated values possibly due to diffraction effects.

The intersection of a 180° wall with a surface parallel to the direction of magnetization gives rise to a narrow strip of free poles with which a finite magnetostatic energy must be associated (Néel 1955, Kaczér 1956). In very small crystals, with dimensions of the same order as the wall-widths, this energy may be comparable with the other terms involved, and the width and energy of the walls may be expected to be modified on this account (Amar 1958). The deposits above 180° walls in powder patterns on Perminvar (Williams and Goertz 1952), silicon–iron crystals (Harrington 1958), and iron whiskers (De Blois and Graham 1958) have been shown to break up on the application of small normal fields, indicating that the sense of the rotation of the magnetization vectors throughout the walls alternates in such a way as to produce short strips of alternate polarity. This internal wall structure reduces the magnetostatic energy in the same way as the formation of domains themselves, but also increases the anisotropy energy associated with the wall (Harrington 1958). The reduction in the magnetostatic energy and the periodicity of the structure have been calculated and found to compare favourably with experimental observations (Shtrikman and Treves 1960 b).

When the figures of the serrated walls on iron whiskers (§ 5.2) are viewed at a grazing angle, the breaks are seen to correspond to very small deviations from linearity ; it has been shown that this leads to the production of free poles over the internal surfaces of the walls which interact with those at the surface, effecting a further reduction in the magnetostatic energy (Shtrikman and Treves 1960 c). The absence of this structure in cobalt or barium ferrite crystals may be due to the narrowness of the 180° wall.

§10. Summary

In the preceding section an attempt has been made to present a wide range of results, although this has precluded any extensive discussion of individual topics. It remains to consider how far these results conform to the general principles laid down in § 2.

The general form of the domain structures in practically all the well-known magnetic materials, in the demagnetized state, is now well established. They have been shown to depend upon the nature of the anisotropy and upon the condition that free poles should either not be formed or that their effect be minimized. The effect of shape of specimens has also been demonstrated and the existence of several postulated structures based on the general principles discussed in § 2 has been confirmed. Structures which owe their existence to stresses in the material, or the distortion of structures by local stresses or structural defects, have been explained in terms of the variation of the magnetic properties. These qualitative studies frequently indicate the domain sizes in particular specimens and form a useful reference for comparison with the structures of relatively new materials, but it is not generally possible to derive quantitative relations for the spacing and energy of the structures observed, because of their complexity and the consequent difficulty in calculating the energy terms involved.

It is only when the domains have a particularly simple structure that reasonable calculations of the energy and domain sizes can be made. In thin crystals of barium ferrite, for example, the measurements give a satisfactory confirmation of the type

of approach outlined in § 2, and the observed sizes of the particles in which domain formation ceases to be favourable correspond well with the calculated values. Although few direct measurements of the properties of domain walls, on which most of the quantitative work is more or less dependent, have been made, such information as has been obtained does not conflict with the basic principles.

Magnetization curves, or magnetic properties in general, can usually be interpreted only in terms of the domain structures involved, and their behaviour in applied fields. Since accurate quantitative accounts of the structures in demagnetized specimens cannot always be given, and a different structure corresponds to each value of the applied field, it is again only possible to relate the domain changes to the magnetization curves in a limited number of cases. Moreover, these usually involve very small specimens on which measurements are difficult, but useful approximations for more complex structures have also been mentioned. While magnetization curves cannot be discussed here at any length, attention may be drawn to the square loops which are connected with the special shape of a specimen or with the discrepancy between the fields required to nucleate domains and to produce saturation ; to the explanation of the constricted loops in alloys sensitive to magnetic heat treatment in terms of localized induced anisotropy or wall anchoring ; and to the suggestion that the square loops of certain ferrites may be associated with extensive inhomogeneities in the grains.

It may also be noted that the nucleation of domains does not form a necessary part of the magnetization cycle as usually measured since extensive surface structures persist in uniaxial, and probably in other crystals, even when they are nominally saturated. Although domain studies have given little direct information on nucleation itself, the observation that domains spread from a limited number of fixed points in barium ferrite crystals after saturation in the easy direction suggests the importance of localized imperfections ; while the very different patterns which are produced after the application of large fields normal to the easy axis indicate that in these conditions nucleation of the domains may occur anywhere within the specimen.

It is also clear from the latter observation that there is not necessarily a single typical structure for a given specimen in the demagnetized state, since this may be achieved in several quite distinct ways. It has been suggested that domains should not form in a uniformly magnetized specimen which has no structural defects (Brown 1945). This difficulty is overcome by considering either demagnetization after saturation in a direction normal to an easy axis, or thermal demagnetization.

Structural defects and the local non-uniformity of demagnetizing fields and anisotropy are probably of great importance in the nucleation of domains in practical specimens, and are the governing factors in several aspects of domain behaviour. The precipitation hardened magnets and elongated single domain powder magnets may be taken as extreme examples of non-uniformity, the magnetic properties and the distribution of the magnetization being so dependent on their structure that the usual principles of domain formation may no longer apply, and such domains as are indicated are probably of a special type.

Finally, it may be noted that although it is suggested that most domain structures and phenomena are fairly well understood, the theoretical approach as a whole is open to criticism. For example, the energy of an array of 180° domains with plane

parallel sides in a crystal with specific dimensions may be shown to be less than that of domains with square section, or of cylindrical structure and so forth, but it has not been found possible to minimize this energy in a purely general way and to show that it is less than that of any other structure in the particular specimen. The calculations apply only to structures which have been demonstrated experimentally or to models which can change in a strictly limited manner, and do not predict actual domain structures. Moreover, it has been pointed out that it is not ' correct ' to calculate the properties of a domain wall by minimizing its energy in two dimensions and then to use these values in a three-dimensional problem. In a really rigorous theoretical approach the presence of walls should not be assumed, and they should appear as part of the general solution for the distribution of the magnetization (Brown 1958). The magnitude of the problem is apparent, and it has not so far been found possible to predict real domain structures although significant results have been produced on the sizes and the precise manner in which the magnetization is reversed in single-domain particles (Brown 1957, 1958, Aharoni and Shtrikman 1958, Frei, Shtrikman and Treves 1957). The general study of the microscopic distribution of magnetization, which has been named ' micromagnetics ', has recently been reviewed (Aharoni 1959, Brown 1959).

Acknowledgments

The authors wish to thank all those who have communicated reprints, unpublished notes, and illustrations and the publishers and authors of all previously published diagrams for permission to reproduce them. They are also very grateful to colleagues who have discussed the present work, and particularly to Professor L. F. Bates for his attention to the manuscript.

References

Abels, V. R., and Sur, J. S., 1956, *Fiz. Metal. Metalloved.*, **3**, 433.
Aharoni, A., 1959, *J. Appl. Phys.*, **30**, 705.
Aharoni, A., and Shtrikman, S., 1958, *Phys. Rev.*, **109**, 1522.
Akulov, N. S., 1929, *Z. Phys.*, **54**, 582.
Amar, H., 1958, *J. Appl. Phys.*, **29**, 989.
Andrä, W., 1956 a, *Ann. Phys., Lpz.*, **17**, 233 ; 1956 b, *Ann. Phys., Lpz.*, **17**, 78 ; 1956 c, *Ann. Phys., Lpz.*, **19**, 10 ; 1959, *Ann. Phys., Lpz.*, **3**, 334.
Bates, L. F., 1959, *Advanc. Sci., Lond.*, **62**, 13.
Bates, L. F., and Carey, R., 1960 a, *Proc. Phys. Soc.*, **76**, 754 ; 1960 b, *Proc. Phys. Soc.*, **75**, 880.
Bates, L. F., and Christoffel, D. A., 1957, *Proc. Inst. Elect. Engrs*, Suppl. No. 5, 231, 245.
Bates, L. F., and Clow, H., 1959, *J. Phys. Radium*, **20**, 93.
Bates, L. F., Clow, H., Craik, D. J., and Griffiths, P. M., 1958, *Proc. Phys. Soc.*, **72**, 224.
Bates, L. F., Craik, D. J., and Griffiths, P. M., 1958, *Proc. Phys. Soc.*, **71**, 789.
Bates, L. F., Craik, D. J., Griffiths, P. M., and Isaac, E. D., 1960, *Proc. Roy. Soc.* A, **253**, 1.
Bates, L. F., and Davis, P. F., 1956, *Proc. Phys. Soc.* B, **69**, 1109.
Bates, L. F., and Hart, A., 1953, *Proc. Phys. Soc.* A, **66**, 813 ; 1956 a, *Proc. Phys. Soc.* B, **69**, 497 ; 1956 b, *Proc. Phys. Soc.* B, **69**, 1200.
Bates, L. F., and Isaac, E. D., 1959, *Proc. Phys. Soc.*, **73**, 14.
Bates, L. F., and Martin, D. H., 1956, *Proc. Phys. Soc.* B, **69**, 145.
Bates, L. F., and Mee, C. D., 1952, *Proc. Phys. Soc.* A, **65**, 129.
Bates, L. F., and Neale, F. E., 1950, *Proc. Phys. Soc.* A, **63**, 374.
Bates, L. F., and Wilson, G. W., 1951, *Proc. Phys. Soc.* A, **64**, 691.

BERTAUT, F., and PAUTHENET, R., 1957, *Proc. Inst. Elect. Engrs*, B, **104**, Suppl. 1, No. 5, 261.
BITTER, F., 1932, *Phys. Rev.*, **41**, 507.
BLACKMAN, M., and GRÜNBAUM, E., 1957, *Proc. Roy. Soc.* A, **241**, 508 ; 1958, *Proc. Roy. Soc.* A, **245**, 408.
BLOOR, D., and MARTIN, D. H., 1959, *Proc. Phys. Soc.*, **73**, 694.
BOZORTH, R. M., 1951, *Ferromagnetism* (New York : Van Nostrand) ; 1954, *Phys. Rev.*, **96**, 311.
BOZORTH, R. M., TILDEN, E. F., and WILLIAMS, A. J., 1955, *Phys. Rev.*, **99**, 1788.
BOZORTH, R. M., and WALKER, J. G., 1950, *Phys. Rev.*, **79**, 888.
BROWN, W. F., 1945, *Rev. Mod. Phys.*, **17**, 15 ; 1957, *Phys. Rev.*, **165**, 1479 ; 1958, *J. Appl. Phys.*, **29**, 470 ; 1959, *J. Appl. Phys.*, **30**, 625.
CAREY, R., 1960, *Proc. Phys. Soc.*, **76**, 567.
COLEMAN, R. V., and SCOTT, G. G., 1958, *J. Appl. Phys.*, **29**, 526.
CRAIK, D. J., 1956, *Proc. Phys. Soc.* B, **69**, 647 ; 1959, *Conference on Magnetism, Sheffield*, 1959 (*Brit. J. Appl. Phys.*, **11**, 307).
CRAIK, D. J., and GRIFFITHS, P. M., 1958, *Brit. J. Appl. Phys.*, **9**, 276 ; 1959, *Proc. Phys. Soc.*, **73**, 1.
CRAIK, D. J., and ISAAC, E. D., 1960 a, *Inter-Service Met. Res. Council*, 2422 ; 1960 b, *Proc. Phys. Soc.*, **76**, 160.
DE BLOIS, R. W., and GRAHAM, C. D., 1958, *J. Appl. Phys.*, **29**, 931.
DILLON, J. F., 1958 a, *J. Appl. Phys.*, **29**, 539 ; 1958 b, *J. Appl. Phys.*, **29**, 1286.
ELMORE, W. C., 1938 a, *Phys. Rev.*, **54**, 1092 ; 1938 b, *Phys. Rev.*, **53**, 757.
ELSCHNER, B., and ANDRÄ, W., 1955, *Fortschr. Phys.*, **3**, 163.
FOWLER, C. A., and FRYER, E. M., 1954 a, *Phys. Rev.*, **94**, 52 ; 1954 b, *Phys. Rev.*, **95**, 564.
FOWLER, C. A., FRYER, E. M., and TREVES, D., 1960, *J. Appl. Phys.*, **31**, 2267.
FREI, E. H., SHTRIKMAN, S., and TREVES, D., 1957, *Phys. Rev.*, **106**, 446.
FULLER, H. W., and HALE, M. E., 1960, *J. Appl. Phys.*, **31**, 238.
GALT, J. K., 1952, *Phys. Rev.*, **85**, 664.
GALT, J. K., ANDRUS, J., and HOPPER, H. G., 1953, *Rev. Mod. Phys.*, **25**, 93.
GELLER, S., and GILLEO, M. A., 1957, *Acta Cryst.*, **10**, 239 ; 1957, *J. Phys. Chem. Solids*, **3**, 30.
GERMER, L. H., 1942, *Phys. Rev.*, **62**, 295.
GOMI, Y., and ODANI, Y., 1960, *J. Phys. Soc. Japan*, **15**, 535.
GOODENOUGH, J. B., 1954, *Phys. Rev.*, **95**, 917 ; 1956, *Phys. Rev.*, **102**, 356.
GRAHAM, C. D., and NEURATH, P. W., 1957, *J. Appl. Phys.*, **28**, 888.
HARADA, T., and TATSUMOTO, E., 1959, *J. Sci. Hiroshima Univ.*, **A23**, 195.
HARRINGTON, J. V., 1958, *Lincoln Lab. Tech. Reps*, 166 (M.I.T.).
HUBER, E. E., SMITH, D. O., and GOODENOUGH, J. B., 1958, *J. Appl. Phys.*, **29**, 294.
ISAAC, E. D., 1959, *Proc. Phys. Soc.*, **74**, 786.
IWATA, T., and YAMAMOTO, M., 1956, *Sci. Rep. Res. Insts Tôhoku Univ.* A, **8**, 293.
KACZÉR, J., 1955, *Czech. J. Phys.*, **5**, 239 ; 1956, *Czech. J. Phys.*, **6**, 310 ; 1957, *Czech. J. Phys.*, **7**, 557 ; 1958, *Czech. J. Phys.*, **8**, 278, and *J. Appl. Phys.*, **29**, 569.
KACZÉR, J., and GEMPERLE, R., 1959, *Czech. J. Phys.*, **9**, 306 ; 1960, *Czech. J. Phys.*, **10**, 505.
KACZÉR, J., GEMPERLE, R., and HAUPTMANN, Z., 1959, *Czech. J. Phys.*, **9**, 606.
KERSTEN, M., 1943, *Grundlagen einer Theorie der Ferromagnetischen Hysterese und der Koerzitivkraft* (Leipzig : Hirzel).
KIRENSKI, L. V., and DEGTIAREV, I. F., 1959, *J. Exp. Theor. Phys., U.S.S.R.*, **35**, 584.
KIRENSKI, L. V., and VETER, V. V., 1959, *Dokl. Akad. Nauk S.S.S.R.*, **125**, 526.
KITTEL, C., 1946, *Phys. Rev.*, **70**, 965; 1949, *Rev. Mod. Phys.*, **21**, 541.
KITTEL, C., and GALT, J. K., 1956, *Solid State Physics*, **3**, 437.
KNOWLES, J. E., 1960, *Proc. Phys. Soc.*, **75**, 885.
KOOY, C., 1957, *Philips Tech. Rev.*, **19**, 286.
KOOY, C., and ENZ, U., 1960, *Philips Res. Rep.*, **15**, 7.
LAWTON, H., and STEWART, K., 1948, *Proc. Phys. Soc.* A, **62**, 193.
LEE, E. W., 1953, *Proc. Phys. Soc.* A, **66**, 623.
LEE, E. W., CALLABY, D. R., and LYNCH, A. C., 1958, *Proc. Phys. Soc.*, **72**, 233.
LIFŞHITZ, E., 1944, *J. Phys. U.S.S.R.*, **8**, 337.
LILLEY, B. A., 1950, *Phil. Mag.*, **41**, 792.

LUBORSKY, F. E., PAINE, T. O., and MENDELSOHN, L. I., 1959, *Powder Metallurgy*, **4**, 57.
MÁLEK, Z., and KAMBERSKY, V., 1958, *Czech. J. Phys.*, **8**, 416.
MARTIN, D. H., 1957, *Proc. Phys. Soc.* B, **70**, 77.
MARTIUS, O. M., GOW, K. V., and CHALMERS, B., 1951, *Phys. Rev.*, **82**, 106.
MARTON, L., SIMPSON, J. A., and LACHENBRACH, S. H., 1954, *J. Res. Nat. Bur. Stand.*, **52**, 97.
MAYER, L., 1957, *J. Appl. Phys.*, **28**, 975.
MENYUK, N., and GOODENOUGH, J. B., 1955, *J. Appl. Phys.*, **26**, 8.
MOON, R. H., 1959, *J. Appl. Phys.*, **30**, 82S.
NÉEL, L., 1944 a, *Cah. Phys.*, **25**, 21 ; 1944 b, *J. Phys. Radium*, **5**, 241 ; 1944 c, *J. Phys. Radium*, **5**, 265 ; 1946, *Ann. Univ. Grenoble*, **22**, 299 ; 1954, *J. Phys. Radium*, **15**, 225 ; 1955, *C.R. Acad. Sci., Paris*, **241**, 533.
NESBITT, E. A., and WILLIAMS, H. J., 1950, *Phys. Rev.*, **80**, 112.
NOSKOV, M. M., and SOKOLOV, A. V., 1947, *J. Expt. Theor. Phys. U.S.S.R.*, **17**, 969.
OATLEY, C. W., and EVERHART, T. E., 1957, *J. Electronics*, **2**, 1.
OSBORN, J. A., 1945, *Phys. Rev.*, **67**, 351.
PAXTON, W. S., and NILAN, T. G., 1955, *J. Appl. Phys.*, **26**, 994.
PRUTTON, M., 1959, *Phil. Mag.*, **4**, 1063.
RHODES, P., and ROWLANDS, G., 1954, *Proc. Leeds Phil. Lit. Soc.*, **6**, 191.
ROBERTS, R. W., and BEAN, C. P., 1954, *Phys. Rev.*, **96**, 1495.
SCOTT, G. G., and COLEMAN, R. V., 1957, *J. Appl. Phys.*, **28**, 1512.
SHERWOOD, R. C., REMEIKA, J. P., and WILLIAMS, H. J., 1959, *J. Appl. Phys.*, **30**, 217.
SHTRIKMAN, S., and TREVES, D., 1960 a, *J. Appl. Phys.*, **31**, 725 ; 1960 b, *J. Appl. Phys.*, **31** ,1475 ; 1960 c, *J. Appl. Phys.*, **31**, 1304.
SMIT, J., and WIJN, H. P. J., 1954, *Advanc. Electronics Electron Phys.*, 6.
SPACEK, L., 1956, *Czech. J. Phys.*, **6**, 256 ; 1959 a, *Czech. J. Phys.*, **9**, 186 ; 1959 b, *Czech. J. Phys.*, **9**, 200.
SPIVAK, G. V., 1957, *J. Acad. Sci. U.S.S.R.*, **21**, 1169.
STEPHAN, W., 1955, *Exp. Tech. Phys.*, **1**, 1.
STEWART, K. H., 1954, *Ferromagnetic Domains* (Cambridge : University Press).
STONER, E. C., 1945, *Phil. Mag.*, **36**, 803 ; 1948, *Rep. Progr. Phys.*, **11**, 43; 1950, *Rep. Progr. Phys.*, **13**, 83.
STONER, E. C., and WOHLFARTH, E. P., 1948, *Phil. Trans. Roy. Soc.* A, **240**, 599.
SUGIURA, I., 1951, *Busseiron Kenkiu*, **38**, 106.
SUR, J. S., 1959, *J. Phys. Radium*, **20**, 113.
SUR, J. S., and ABELS, V. R., 1955 a, *Fiz. Metal. Metalloved.*, **1**, 11 ; 1955 b, *Dokl. Acad. Nauk S.S.S.R.*, **102**, 449; 1955 c, *Dokl. Akad. Nauk S.S.S.R.*, **104**, 209 ; 1955 d, *Dokl. Acad. Nauk S.S.S.R.*, **105**, 469.
SUR, J. S., STOL'C, E. V., KANDAUROVA, G. S., and BULATOVA, L. V., 1957, *Fiz. Metal. Metalloved.*, **5**, 234.
SUR, J. S., and ZAJKOVA, V. A., 1958 a, *Fiz. Metal. Metalloved.*, **6**, 545 ; 1958 b, *Fiz. Metal. Metalloved.*, **6**, 647.
TANIGUCHI, S., and YAMAMOTO, M., 1954, *Sci. Rep. Res. Insts Tôhoku Univ.* A, **6**, 330.
WILLIAMS, H. J., 1952, *Bell Lab. Rec.*, **30**, 385.
WILLIAMS, H. J., BOZORTH, R. M., and SHOCKLEY, W., 1949, *Phys. Rev.*, **75**, 155.
WILLIAMS, H. J., and GOERTZ, M., 1952, *J. Appl. Phys.*, **23**, 316.
WILLIAMS, H. J., and SHERWOOD, R. C., 1957, *J. Appl. Phys.*, **28**, 548 ; 1958, *Trans. Amer. Soc. Metals*, 35.
WILLIAMS, H. J., SHERWOOD, R. C., and SHOCKLEY, V., 1949, *Phys. Rev.*, **75**, 178.
WILLIAMS, H. J., and SHOCKLEY, W., 1949, *Phys. Rev.*, **75**, 178.
WILLIAMS, H. J., and WALKER, J. G., 1951, *Phys. Rev.*, **83**, 634.
WOHLFARTH, E. P., 1959, *Phil. Mag.*, **8**, 87.
YAMAMOTO, M., and IWATA, T., 1950, *Phys. Rev.*, **81**, 887 ; 1953, *Sci. Rep. Res. Insts Tôhoku Univ.* A, **5**, 433.
YAMAMOTO, M., and TANIGUCHI, S., 1956, *Sci. Rep. Res. Insts Tôhoku Univ.* A, **8**, 280.
YAMAMOTO, M., TANIGUCHI, S., and AOYAGI, K., 1956, *Phys. Rev.*, **102**, 1295 ; 1958, *Sci. Rep. Res. Insts Tôhoku Univ.* A, **10**, 20.

PHOTOELECTRONIC IMAGE INTENSIFIERS

By J. D. McGEE

Imperial College, London

CONTENTS

	PAGE
§ 1. Introduction : Statement of problem and its relation to photography	168
§ 2. Single-stage image intensifiers	172
2.1. Image intensifier with phosphor screen and transfer optics	172
2.2. Image intensifier with thin window, phosphor screen and contact recording	174
2.3. Lime-soda glass, thin-window image intensifier	177
2.4. Electron-transmitting mica-window image tubes	181
2.5. Lallemand's electronic camera	182
2.6. The two-chamber, single-stage image-recording tube	186
2.7. Electrostatic focusing, single-stage tube	187
2.8. X-ray image intensifier	188
2.9. High-speed shutter image tube	189
§ 3. Multi-stage image intensifiers	190
3.1. General considerations	190
3.2. Cascaded phosphor–photocathode intensifying screen tubes	191
3.3. Transmission secondary-emission image intensifier	196
3.4. The channelled secondary-emission image multiplier	200
§ 4. Positive feedback image intensifying systems	203
4.1. General considerations	203
4.2. Single-tube system with accurate registration	204
4.3. Systems without registration	205
§ 5. Image intensifier tube applications	205
5.1. Astronomy	205
5.2. Nuclear physics	207
5.3. Spectroscopy	208
5.4. Radiology	209
Acknowledgments	210
References	210

Abstract. The limits set to the detectability of an image by the statistical fluctuations resulting from the quantum nature of light are analysed and the efficiency of the photoelectric effect as a detector of optical images is compared with that of the photographic effect.

The possible methods of using the photoelectric effect for image detection are enumerated and it is pointed out that this review deals only with those image intensifiers using free electrons.

The second section outlines the possible light gain to be had from a simple photocathode–phosphor image tube and gives the reasons for its failure to reach the ideal performance. The methods employed to enable these tubes to be operated at high gain without undue background and to enable the output light to be used efficiently are described. The techniques developed allow the electron image to be projected through a thin window on to a fine grained photographic film external to the tube. The big advantages of this technique are pointed out.

The method developed by Lallemand over the years in which the photocathode and photographic emulsion are in the same vacuum chamber is described and its advantages and disadvantages are enumerated. Then the development of this method is described in which the electron image is projected through a thin film separating the photocathode and photographic plate compartments, the thin film protecting the photocathode from the gas evolved from the photographic emulsion.

A single-stage electrostatic focusing image intensifier in which electron-image demagnification is used is described and its operation discussed. The elaboration of this type of image intensifier as an x-ray image intensifier is described.

The methods of electron-image multiplication are described in the third section. In these the number of electrons is multiplied by a large factor before they are accelerated and detected. Three methods are described: firstly, the cascaded phosphor–photocathode screen intensifier, secondly, the transmission secondary-emission multiplying screen method, and thirdly, the channelled electron-image multiplier. The problems involved in each of these devices are outlined and the present state of development and performance is summarized.

Section 4 deals with methods of positive light feedback from the output image of a simple image tube to the photocathode. The fundamental requirements for this method to work are explained and the methods employing exact registration and non-registration are discussed.

In the last section, brief descriptions are given of scientific observational problems in astronomy, nuclear physics, spectroscopy and radiology, in which one or other of the image intensifier tubes described might find useful application.

§1. Introduction: Statement of Problem and its Relation to Photography

Light images are recorded by the action of the individual photons collected by an optical system and formed into an image on a light-sensitive surface. The action of the photons may be photochemical, photographic or photoelectric, though the distinction between these effects is more superficial than real.

The ideal situation of maximum efficiency is reached when each photon produces an individual, identical and identifiable record, such as one developed grain in a photographic emulsion, and in these circumstances the quality of the recorded image, which may be defined by its signal/noise ratio, will be determined by the statistical fluctuations of these recorded events. It is of interest to calculate the approximate level of illumination of an average scene at which an intelligible image may be recorded, using a quite conventional optical system and assuming various efficiencies for the recording of individual photons.

If we allow an illumination of I_0 lumen ft^{-2} to fall on a scene comprising features ranging from perfect white to absolute black, the illumination in the image plane I_i corresponding to the areas of peak white is given by $I_i = I_0 T/4F^2$, where F is the reciprocal of the relative aperture of the lens and T is its transmission efficiency for the light used. If the image covers a useful area of A ft^2 and we assume it to have an aspect ratio a and to consist of N television lines, or N^2a picture points, that the exposure is for t sec and that one lumen is equivalent to P photons per second, then the number of photons arriving on a picture point in a peak white area in time t is

$$n = \frac{I_0 TAPt}{4F^2N^2a}. \qquad \ldots\ldots(1)$$

If it is assumed that the arrival of photons in the image is random in space and time, and that the probability of detection is ϵ, then in an area of nominally uniform brightness B, corresponding to a mean accumulation of photons of n per picture point, there will be a mean deviation from this average value of $\sqrt{n}$, and it will not be easy to discriminate between the recorded impression of two adjacent areas of nearly the same brightness B and $B+\Delta B$, which have a mean difference of $N(2n\epsilon)$

unless $\Delta B/B > \sqrt{(2n\epsilon)}/n\epsilon$. It is usual to assume a factor k by which the brightness step must be greater than the mean deviation, or ' r.m.s. noise ', in order to make it possible to recognize the brightness step with a given degree of certainty. This factor k is usually termed the coefficient of certainty.

Thus, the minimum contrast difference recognizable with a given degree of certainty is given by

$$\left(\frac{\Delta B}{B}\right) \geqslant k\left(\frac{8F^2N^2a}{I_0TAPt\epsilon}\right)^{1/2} \qquad \ldots\ldots(2)$$

or where the peak white signal/r.m.s. noise ratio

$$S = n\epsilon/\sqrt{(n\epsilon)} = \sqrt{(n\epsilon)}$$
$$= \left(\frac{I_0TAPt\epsilon}{4F^2N^2a}\right)^{1/2}. \qquad \ldots\ldots(3)$$

Thus we see that the minimum contrast step detectable becomes smaller with increase in the total number of photons integrated in the picture. It also decreases with increase in the area of the picture elements we wish to observe. If the area of the picture element to be distinguished is small, of the order of one picture point in a 500-line picture, the value of k to ensure reliable detection is approximately 5. However, for an extended image from which the eye has much ordered information to guide it in recognizing the pattern through the random noise, a value of k nearer to one can be tolerated. Thus, if the peak white signal is ten times the r.m.s. noise arising in the same area from the shot noise of the photoemission, it is known from television experience that the picture will be intelligible though ' noisy '. As the value of S decreases further the information content of the picture decreases rapidly.

If now we assume a value of $S = 10$ and evaluate the expression (3) for easily realizable values of the parameters involved, such as $T = 0{\cdot}8$, $A = 1/36$ ft^2, $a = 1$, $F = 1{\cdot}5$, $N = 500$, $t = 1/10$ sec, $\epsilon = 1$, $P = 1{\cdot}3 \times 10^{16}$ photons/sec (the figure for white light), then $I_0 \sim 10^{-5}$ lumen ft^{-2}.

This is about one order of magnitude less than the ambient illumination on a clear star-lit night ($\sim 10^{-4}$ lumen ft^{-2}) and is about three orders of magnitude lower than the minimum illumination required to obtain this signal/noise ratio from the most sensitive operational television camera ($\lesssim 1/10$ lumen ft^{-2}) and the discrepancy is one or two more orders of magnitude greater for photography. This suggests that $\epsilon \ll 1$ in both these methods of recording.

The failure to reach the ideal picture quality is due partly to the fact that other sources of noise, such as amplifier noise in the case of a television picture, dark current in an image tube, or emulsion fog in the case of a photograph, degrade the image. However, it is also largely due to the fact that only a small percentage of the incident photons produce a definite recordable event such as the activation of a silver halide grain or the ejection of a photoelectron. Lumping all sources of noise together, we obtain a certain signal/noise ratio which is the same as would be obtained if the noise were due only to the genuine shot noise of the recording of photons with an efficiency of $\alpha\%$. Then α is the ' equivalent quantum efficiency ' (Rose 1948) of the system employed. However, the real quantum efficiency is much larger than this, and if the incidental sources of noise could be eliminated greater efficiency of recording could be achieved.

It has been well known for some years that the efficiency of conversion of photons to electrons by an efficient modern photocathode (i.e. the photoelectric quantum efficiency) is about 10%, i.e. about two orders of magnitude greater than the corresponding figure for the efficiency of activation of silver halide grains by photons incident on photographic emulsion, which is about 0·1%. Thus the information inherent in an optical image of a given exposure that can be preserved in a photoelectronic image formed from it can be greater, by this same factor, than

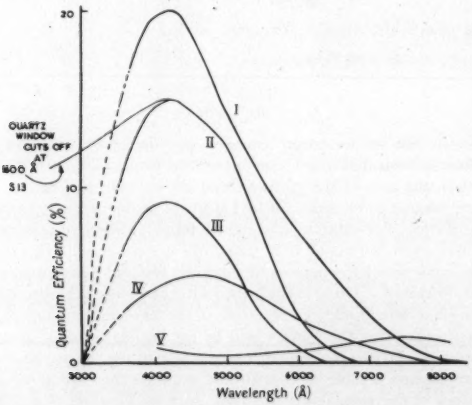


Fig. 1. Photoelectric quantum efficiency curves of principal photocathodes.

that in the corresponding photographic record. Provided the electron image can be recorded without loss of information, the photoelectric technique should be about one hundred times as efficient as the photographic method of recording.

There are three other advantages of the photoelectric effect as a means of detecting light. Firstly, the photoelectric response to light is strictly linear over all intensities whereas the photographic response is far from linear even in the normal working range of exposure. Secondly, when the light flux is reduced to very low levels—and it is these levels that present the real problems in modern observation—the photographic reciprocity law breaks down and the emulsions become progressively less efficient. Thirdly, the photoelectric cathodes respond efficiently to light over a wide range of wavelengths. The most versatile photocathode covers a greater range of wavelengths ($\lambda = 2000 \rightarrow 8500$ Å) than any single photographic emulsion and in the infra-red the best photocathode is more efficient and responds to longer wavelengths than the best photographic emulsion.

The response curves giving the quantum efficiency at different wavelengths for several of the more important photocathodes are shown in Fig. 1. They all stretch

as far into the ultra-violet as measurements have been made with very little fall in efficiency and the practical limit is set by the transparency of the envelope, since all these surfaces must be protected from the air. The most versatile is probably the tri-alkali cathode (S-20) (Sommer 1955) which has a response shown as curve I, stretching from a long wave threshold at about 8500 Å, to a peak at about 4500 Å, and then stretching on into the ultra-violet as far as the transparency of the envelope will allow. The response of the better known antimony–caesium cathode (S-11) ($SbCs_3$), prepared with oxide substrate, is shown in curve II. This has a shorter threshold wavelength ($\sim$6500 Å) but is almost as efficient in the blue and ultra-violet. It has a low dark current. Curve III shows the response of an antimony–caesium photocathode (S-9) prepared without oxidation. While this has a lower sensitivity it also has a shorter threshold wavelength and a very low dark current. Curve IV is the response curve of the bismuth–silver–oxygen–caesium (S-10) cathode which was useful for its efficient red response before the advent of, and is still a competitor with, the tri-alkali cathode. Finally, curve V is the response of the silver–caesium–oxide–silver–caesium cathode (S-1), which stretches the furthest of any into the infra-red ($\sim$12 000 Å). Thus, from 7500 Å to 12 000 Å it is the most efficient photocathode but it is much less efficient than the other four in the visible range. It suffers from the disadvantage that its natural thermal emission at normal temperatures is high—probably three orders of magnitude greater than that from an antimony–caesium cathode.

These characteristics of the photoemissive effect offer, in principle, great advantages in the recording of light, and it is the problem of effectively recording electron images produced by optical images in such a way as to take full advantage of this large potential gain that forms the subject of this paper.

Photoelectronic image detectors may be divided into three classes as follows :

(i) Those in which the photons of an optical image are converted into free electrons which may be given energy and increased in number and formed into an electron image which can be recorded or observed directly.

(ii) Those in which the electrons excited by the photons of an optical image travel within a solid where they may be multiplied and given energy to produce a fluorescent image.

(iii) Those in which the electrons of the image derived from an optical image are used to produce a charge image on a two-dimensional array of minute, parallel capacitors, which are discharged by scanning with a beam of electrons to produce a television picture signal, which can be recorded or used to reproduce a television reproduction of the original optical image.

These techniques are used for three main purposes : (*a*) To intensify the original optical image. That is, to achieve a greater effective output of light as compared with the light input of the same wavelength. (*b*) To convert the input radiation to light of a different wavelength ; e.g. infra-red, x-rays or ultra-violet to visible light. (*c*) To transmit an image over a distance by cable or radio link. Only devices of class (iii) above can be used for this purpose though the same type of device may achieve also the purposes of (*a*) and (*b*).

In this article we shall deal only with those devices of class (i) which are generally known as image intensifiers.

§ 2. Single-stage Image Intensifiers

These are defined as tubes in which there is only one generation of free electrons.

2.1. *Image Intensifier with Phosphor Screen and Transfer Optics*

The simplest and most obvious image intensifier is that in which the photons of an optical image liberate, from a photocathode, free electrons which are accelerated and focused to form an electron image on a fluorescent screen where the electron energy produces a fluorescent light-image. Such a device is illustrated in Fig. 2. It consists of a highly evacuated glass cylinder 1 with flat windows, 2 and 3, at either end, on which are formed respectively a transparent, conducting photocathode and

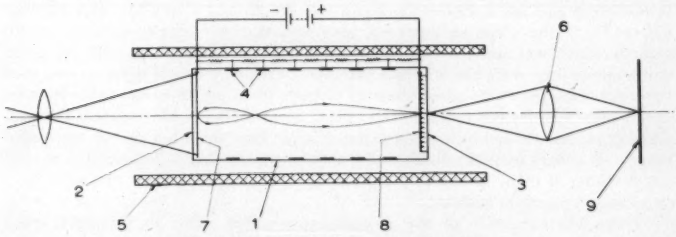


Fig. 2. Simple image converter or intensifier.

a fluorescent screen. A fairly high potential difference, of 5 to 25 kv, is maintained between the cathode and the screen (anode) and the field gradient between them is kept fairly uniform by a series of conducting ring electrodes 4 formed on or near the tube walls and held at progressively increasing potentials. The tube is surrounded by a solenoid 5, through which a steady current is passed to maintain a fairly uniform axial magnetic field throughout the tube. As is well known, with this arrangement of uniform axial electric and magnetic fields, slow photoelectrons leaving the photocathode may be brought to a ' focus ' at the fluorescent screen after having performed an integral number of cycloidal loops in the magnetic field between cathode and screen. The relationship between electric potential difference V, magnetic field B, and the distance d between cathode and anode for focus is given by the expression

$$d = nK\sqrt{V}/B \qquad \ldots\ldots(4)$$

where K is a constant and n is the number of cycloidal loops in the electron path.

It is clear that this device gives us a means of injecting energy into the system by accelerating the electrons, and if the photocathode and fluorescent screen efficiencies are high enough and a sufficiently high accelerating voltage can be applied, more light may be liberated by the photoelectrons from the fluorescent screen than was required to fall on the photocathode to liberate them (Mandel 1955).

For a good quality Sb–Cs–O (or S-11) photocathode the average quantum efficiency for white light is about 5% (Engstrom 1960); that is, about one photon in 20, on average, liberates one photoelectron. If the liberated photoelectrons are accelerated to, say, 20 kev energy and impinge on a phosphor screen, which we

will assume to be the most efficient (20%) blue ZnS : Ag (Bril and Klasens 1952), which gives off light of an average wavelength of about 4500 Å, and corresponding to a photon energy of approximately 2·75 ev, then we shall have $(20\,000/2{\cdot}75) \times 0{\cdot}20 = 1450$ photons liberated by each electron incident on the phosphor screen.

The photons are radiated fairly isotropically and only about 35% emerge directly from the required side of the screen. Some 35% are lost by absorption in the screen or glass and another 30% are reflected from the thin, but opaque and highly reflecting aluminium film which lies over the surface of the phosphor upon which the photoelectrons impinge. The radiation from the screen in the forward direction is approximately Lambertian. The aluminium backing film serves several purposes. As we have just noted it reflects a substantial proportion of the emergent photons in the direction of the recording lens and hence increases the percentage of the photons emergent in the forward 2π solid angle from approximately 35% to approximately 65%. Because the phosphor layer is quite thin, e.g. less than 10 μ, this reflected light does not seriously reduce image definition or contrast until the dimensions of the image detail become comparable with this thickness. The aluminium film also prevents light from being radiated back towards the photocathode where it would produce more electrons which would be accelerated to the screen and cause loss of contrast. This feedback from screen to cathode can become cumulative, and result in an unstable state being reached. It was, in fact, only after the introduction of the aluminium backing that efficient image intensifiers became practicable. Again, the aluminium film stabilizes the potential of the phosphor screen firmly at anode potential. In its absence the screen potential must be maintained by secondary electron emission, and in many cases it may be substantially lower than that of the anode. Lastly, the aluminium film forms a useful protecting barrier against poisoning of the phosphor screen, for example by caesium vapour during the activation of a photocathode in the same vacuum envelope.

Now of the approximately 1500 photons liberated in the phosphor by each incident photoelectron about 65%, or about 1000, will emerge in the forward 2π solid angle. As 20 photons were required to produce this one electron the light gain is a factor of 50. However, to use these emergent photons, either photographically or visually, an optical system 6 (Fig. 2) must be used to form an image on a photographic plate 7 (Fig. 2) of the image on the phosphor. Even with an elaborate but conventional optical system it is difficult to transfer more than 10% of the emergent light to a photographic plate and the figure is usually nearer 5%. Thus the overall gain of such a system is not likely to be greater than about 4 times. Besides this poor result there is also the fact that it has been difficult in the past to operate such image tubes at overall accelerating voltages above, say, 10 or 15 kv without unwanted emission of spurious electrons occurring from the tube electrodes. These reached the phosphor screen in sufficient numbers and with sufficient energy to produce an irregular, diffuse glow which was often brighter than the faint images which it was intended to intensify.

Additional defects of such tubes were : an impairment of the geometrical accuracy of the image and, to a greater or less extent, a loss of image definition. In view of all these unsatisfactory features it is not very surprising that this simple

form of image intensifier was not a general success. Only in one case were single-stage image tubes widely used. That was in the conversion of infra-red to visible images for military purposes, and in this case the immediate conversion of an invisible to a visible image outweighed the other defects of the tubes.

2.2. *Image Intensifier with Thin Window, Phosphor Screen and Contact Recording*

For some years one research project in the author's laboratory has been aimed at eliminating the defects of the simple tube outlined in the last section. An earlier report on this work by Zacharov and Dowden (1960, see also McGee 1959)

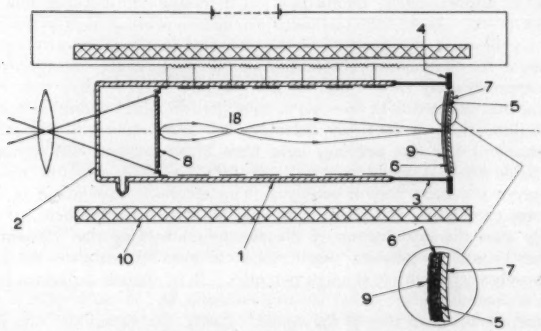


Fig. 3. Thin mica-window image intensifier.

showed that considerable progress had been made towards eliminating the two major faults of the simple tube: the loss of light at the output due to the conventional optical system and the spurious background.

The methods employed will be illustrated by reference to Fig. 3. The tube consisted of a cylindrical Pyrex glass envelope 1 with a flat glass window 2 at one end. At the other end it was joined through a graded-glass section to a metal cylinder 3 which in turn was brazed to a chrome-iron annular ring 4 of 1 in. internal diameter. A sheet of mica 5, between 10 and 15 μ thick, was sealed with a suitable solder glass† between two chrome-iron rings to make a strong and vacuum-tight joint. The gradation of the coefficient of expansion of materials from Pyrex glass ($\sim 35 \times 10^{-7}$) to chrome-iron ($\sim 90 \times 10^{-7}$) is necessary in order to have the end plate material to match the coefficient of expansion of mica. The mica window so formed, 1 in. diameter and approximately 12 μ thick, is remarkably strong, withstanding not only atmospheric pressure but often up to 6 or 7 atmospheres pressure without fracturing. They are vacuum tight and stand the normal vacuum processing bake to 300 or 400°C.

A fluorescent screen 6 was settled on the inner surface of the mica window by more or less conventional methods and an aluminium backing film 9, about

† B.T.H. glass type C.93.

1000 Å thick, was provided. The phosphor used was blue ZnS : Ag, or P.11, of particle size about 3 μ and a layer thickness of 10 μ or less.

After evacuation of the tube the mica window dished inward under atmospheric pressure with a radius of curvature of about 10 cm and, to record a fluorescent image formed on the phosphor, a flexible photographic film 7 was pressed evenly against the outside surface of the mica. This can be done safely, and a fair pressure exerted, because of the remarkable strength of the mica window. Under these conditions all the light emergent from the mica window must fall on the emulsion and hence be effective photographically. However, owing to the finite thicknesses of the phosphor layer and the mica sheet there will be a lateral spread of the light resulting in a loss of definition in the recorded image as compared with the fluorescent image.

It has been known (McGee 1956, Hiltner 1956) for a long time that the origin of spurious background is largely due to the lowering of the work function of the surfaces of the electrodes in the tube by the action of caesium vapour which is admitted to the tube at a fairly high temperature to activate the photocathode. Thus, according to Hiltner (McGee and Lubszynski 1939) reduction in the work function of a surface from 3 to 2 v can result in an increase in field emission by a factor of 10^{14} in a field gradient of 5×10^6 v cm^{-1}. Hence, when high electrical field gradients are applied between the electrodes, emission of electrons occurs from any sharp points or rough edges of the negative electrodes.

To avoid this trouble the design shown in Fig. 3 was adopted (McGee 1956). In this tube the glass plate 8, on which the photocathode was to be formed, was mounted on a shelf some 1½ in. from the end window of the tube. This glass plate was held on the shelf by spring catches which could be operated from outside the tube to release the plate and refix it onto the shelf. The plate and shelf formed a barrier across the tube which, though not gas-tight, was an effective barrier to the passage of caesium vapour. After evacuation and bake of the tube, antimony was deposited from an evaporator, which was inserted through a side tube, onto the surface of the glass plate 8 facing the small end-compartment of the tube. Caesium was then admitted to this compartment and the tube baked to activate the caesium layer. It was found that during this process a negligible amount of caesium vapour passed the barrier and hence the larger, working section 18, of the tube remained free from caesium contamination. After the tube had been sealed off from the vacuum pumps, the glass cathode-plate 8 was released from its retaining catches, turned over, and refixed in position with the transparent photocathode layer now facing into the main section of the tube.

On applying the accelerating potentials of equal steps to the ring electrodes on the walls of this tube it was found that the overall potential at which it could be operated, without visible spurious background appearing on the phosphor screen, was considerably higher than that possible with a similar tube in which caesium had been admitted to the body of the tube in the conventional method of activation.

When this reversible cathode technique was first applied in a tube it was feared that the photocathode might lose its sensitivity because it is exposed in a space in which all surfaces are free from caesium, while in normal photocell technique the photocathode exists in an environment saturated with caesium. It was thought

possible that the caesium would slowly diffuse away from the cathode which would consequently lose its efficiency. Extensive experiments designed specifically to test this possibility have shown that the fears were unfounded. The thin layer of intermetallic compound of antimony and caesium ($SbCs_3$) covered by a few monolayers of pure caesium, which form an efficient photocathode, appears to be quite stable at temperatures up to at least 50°C. Provided the tube vacuum is good these cathodes have remained constant in efficiency in caesium-free environments for long periods.

The tube was operated in the solenoid 10, Fig. 3, which produced a fairly uniform axial magnetic field of about 100 gauss over the space between the cathode and anode. The performance that could be achieved, though not consistently, was as follows: (i) *Background.* This was invisible to the dark-adapted eye when operating at 20 kv, and no detectable blackening appeared on a fast film placed against the window and exposed for 15 min. (ii) *Resolution.* This was 60 line-pairs per mm viewed on the phosphor screen and 30 line-pairs per mm were resolved on a photographic contact print made by pressing a film against the mica window. This resolution was fairly uniform over the whole usable area of the window—about 0·9 in. diameter. The geometrical accuracy of reproduction was fairly good but showed some S-distortion and in some cases general asymmetrical distortion of unknown origin. (iii) *Light gain.* This was measured by using light of the same composition—i.e. light from an excited ZnS:Ag phosphor screen—to illuminate the photocathode as is emitted from the phosphor screen. The gain was measured in two ways. Firstly, the input and output light fluxes were measured by means of a photocell, the ratio of the measurements giving the light gain of the tube. Secondly, the input light was allowed to fall on a photographic emulsion to give a certain density. A series of exposures of the same emulsion was then made of the intensified image by contact with the mica window. The inverse ratio of the times required to obtain the same density of the emulsion gives a measure of the light gain of the tube.

There were considerable discrepancies between the results obtained by these two methods. The light gain factors as measured by the photographic method were larger, by factors varying from 50 to 200%, than those obtained by the photocell measurements. The gain, as was to be expected, was closely proportional to the photocathode sensitivity which varied considerably from tube to tube. A result which was considered good, but attainable in a fair proportion of tubes which did not suffer from vacuum troubles, was 40 μA/lumen, giving a photoelectric gain of 30 times and a photographic gain of 50 times, measured at 20 kv operating potential.

The discrepancy between the photographic and photoelectric measurements of gain has not been resolved. It is possible that it is an instrumental error, but some of the discrepancies were too large for this explanation to be acceptable. The possibility that soft X-rays, produced at the phosphor screen, penetrated the thin mica window and added their effect to that of the visible light in exposing the film while not affecting the photocell, was also examined but with negative results. A very thin opaque layer of, for example, aluminium, which should have allowed soft X-rays, but not light, to pass, was found to stop all effect on the emulsion.

There remains the possibility that the greater gain of the photographic measurement may be real and due to the detailed nature of the photographic process

(Mandel 1959) in which, it is generally accepted, a grain of silver halide must accept at least two photons within a quite short period of time to be rendered permanently developable. If the time interval elapsing between the arrival of two photons is too long, the effect of the first wears off before the second arrives and the halide grain remains unactivated. This accounts for the reciprocity failure of the photographic process at very low light levels. Now in this image tube the photons arrive on the photocathode, or the photographic film used for measuring the light, randomly spaced in time and space. At the phosphor screen end, not only are there about fifty times as many photons but they are liberated in bunches by individual photoelectrons thus concentrating the available light onto small areas for very short times. Thus reciprocity failure would begin to be appreciable for direct photography of the original image before it would affect the result from the much brighter intensified image. However, further careful experiments are necessary to clarify this discrepancy.

These experiments showed fairly conclusively that a tube of this type could be made to give an acceptable performance in respect of the three main requirements : light gain, resolution and background. At the same time it was shown that it was quite practicable to take contact photographs by pressing film against the outside surface of a thin mica window. Large numbers of such photographs were made without failure of the mica windows. Also it was found that film could be pressed against the mica window to be in good contact over the whole surface without difficulty. The amount of geometrical inaccuracy introduced by the slight ' dishing ' of the film as it is pressed against the mica is very small—much less than that due to the electron-optical distortion of the image. To avoid distortion during development of the film it should be cemented back on to a glass plate before processing.

Unfortunately the tubes made to the design described above gave much trouble due to poor vacuum during processing and seldom lasted more than a few days without serious deterioration in cathode sensitivity and general performance. To overcome this fatal limitation a redesign of the tube was undertaken which is described in the next section.

2.3. *Lime-soda Glass, Thin-window Image Intensifier*

The vacuum troubles encountered in making the tube described in the last section were believed to be due to the metal section of the envelope which it was difficult to degas sufficiently thoroughly to ensure an adequate degree of vacuum for a high voltage photoelectric tube. This metal end was introduced solely to provide a chrome-iron end-plate to which the mica window could be sealed since they have approximately the same coefficients of expansion. However, lime-soda glass also has very nearly the same coefficient of expansion and it was easily shown that a mica sheet could equally easily be sealed between two lime-soda glass annuli.

If this were adopted it became necessary either to make the whole tube of lime-soda glass or to insert a graded-glass seal between the lime-soda end-plate and the rest of the tube of Pyrex glass. Of these two alternatives it was decided that the glass-working hazards of the former, though troublesome, were less than those associated with the use of the graded seal.

A long series of experiments carried out by B. Wheeler, B. Zacharov, G. Horvath and the author resulted in the tube shown in Fig. 4. The tube envelope, except the window, is lime-soda glass ; it is a cylinder 1 of about 45 mm internal diameter closed at one end by a flat glass plate 4 and at the other by the mica window 2. It is divided into two sections, the processing chamber 5 and the operating chamber 18, by a close fitting glass annulus 3 and the glass plate 8. Two side tubes 6 and 7 are provided, the latter for the evacuation and introduction of an antimony evaporator to form the basic photocathode layer of antimony on to the glass plate 8

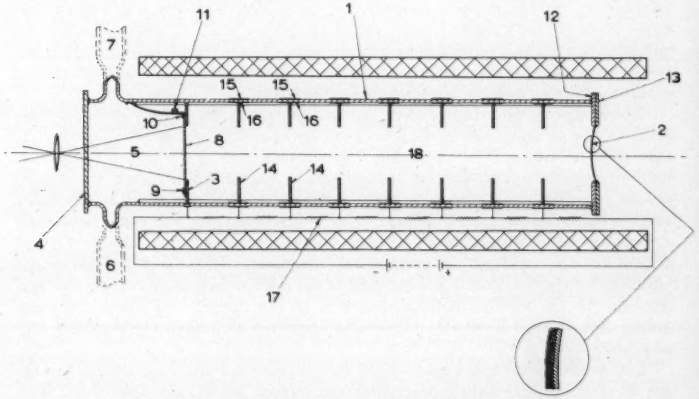


Fig. 4. Lime-soda glass, mica-window, single-stage intensifier.

and the former for the introduction of caesium to activate the antimony layer to form the photocathode. As before, the plate 8 is held in catches 9 and 10, the former of which is fixed and the latter is spring loaded so as to hold the plate firmly on the shelf 3, or slightly ajar, and is operated by the action of a strong magnet brought up outside the tube to attract the iron slug 11 and hence retract the spring and so release the plate 8.

During the main evacuation and degassing bake the plate 8 is held away from its shelf so that the whole tube can be pumped through the one exhaust tubulation 7. The plate can then be dropped onto the shelf to close the opening between the two sections of the tube prior to the beginning of the cathode activation process.

The annular glass plates 12 and 13 are sealed with suitable solder glass (C.93) to the mica sheet which is held as in a sandwich between them. Several short radial strips of platinum were painted on the inner surface of the glass annulus to connect the aluminium backing of the screen through the glass wall to the anode voltage supply. Plate 12 is sealed to the cylinder 1 with solder glass. Provided care is taken in heating and cooling the lime-soda glass these operations can be carried out without serious hazard.

The phosphor screen is deposited on the inner surface of the mica window and the aluminium backing is applied before the end window 4 is sealed on, so the necessary operations can be performed through the open end of the tube. Because there is no metal in the immediate neighbourhood of the phosphor screen in this design, a potent source of 'poisoning' of the phosphor was removed which was a serious trouble in the previous design.

It is desirable to establish between the cathode and the anode a uniform accelerating electric field. An approximation to this is usually achieved by applying a series of conducting rings to the inner walls of the glass tubes, as in Fig. 3, by painting on a colloidal suspension of metal (for example platinum) and then baking to form a bright conducting metal layer. Contact to these rings is made by metal seals through the glass walls of the tube. These rings, however, result in sharp discontinuities in the field gradient at their edges and from these edges electron emission tends to take place. Consideration was given to ways of achieving as nearly as possible a uniform potential gradient along the glass walls. A high resistance Aquadag spiral was applied to the inside wall of the tube but electrical breakdown between adjacent turns occurred in vacuum at high voltages and it became necessary to cover the spiral with an insulating layer to prevent these sparks. It then appeared possible to use the slightly conducting lime-soda glass walls as the spark-suppressing layer and to apply the high resistance coating to the external surfaces of the tube. Since the electron current from cathode to anode in operating conditions is small, probably less than 10^{-7}A, and the photoelectrons are constrained by the axial magnetic field to move in cycloidal paths of quite small diameter, 1 mm or less, it was reckoned that few of them would reach the tube walls. Also as the vacuum must be good, 10^{-7} torr or less, the mean free path of electrons must be very great and very few positive ions which might reach the walls will be formed. Hence in operation the electric currents to the wall electrodes must be minute and it was considered that the electrical leakage through the walls from the external potential-stabilizing electrodes should maintain the correct potential gradient along the internal walls of the tube. This appeared to be so in practice but an unexpected snag arose. When only an electrical potential of 20 kv was applied to a tube negligible background appeared on the phosphor screen, but when an axial magnetic field was also applied, a strong, more or less uniform, glow appeared on the screen and this persisted at quite low voltages, such as 5 kv. The curious facts were that *both* magnetic and electric fields were essential to produce this background and it appeared irrespective of whether the cathode was a photosensitive surface or a clean metal plate.

A tentative hypothesis designed to explain this phenomenon is as follows: in a simple tube as described, a few electrons will be liberated from the walls, by cosmic rays, thermal emission, etc. An axial electric field will, in the absence of a magnetic field, accelerate these electrons towards the phosphor screen and they will also move slightly towards the tube axis. There will be probably only a small number of them and though they do reach the screen with considerable energy they will produce negligible background. If, however, an axial magnetic field is applied, the component of the velocity of such an electron normal to the surface of the tube wall will result in the electron being constrained to move in a cycloidal path which will bring it back to the tube wall, but only after it has moved a considerable

distance parallel to the tube wall in the electric field. Hence it will strike the tube wall with considerable energy, probably several hundred electron volts, which will be enough to cause secondary-electron emission. The lime-soda glass has a fairly high secondary-electron emission coefficient and conditions of magnetic field and voltage gradient are easily arranged to give a secondary-emission ratio much greater than unity. Thus we appear to have an electron-multiplying mechanism along the wall of the tube which might result in a greatly intensified stream of electrons reaching the end of the tube. However, as the emission velocities of secondary electrons are not very high the diameters of their cycloidal paths cannot be very large, for example, not greater than 1 cm. Hence, most of these electrons should arrive round the edge of the screen with comparatively few in the centre. But we found that the screen illumination was fairly uniform and hence the hypothesis is inadequate. It was then necessary to postulate a further effect: namely, that the electrons which are turned back and strike the tube wall with considerable energy also liberate ionized sodium atoms from the surface. Evidence of such an effect has been given by Holland (1959) and Ennos (1953). If this does in fact happen, the liberated ions will move back towards the cathode in cycloidal paths which have diameters and loop lengths about 200 times greater than those of electrons. Hence these ions will move much further away from the wall towards the centre of the tube, some will strike the walls with greater energy but others can impinge on the whole area of the cathode from which they will liberate electrons. These electrons will be accelerated and directed by the magnetic field to corresponding areas of the phosphor screen. Thus we have a mechanism that can account for the observed effects.

If the above hypothesis is correct the remedy for the trouble becomes obvious: baffles must be provided along the walls to suppress the electron avalanche and if possible the returning ion stream. Such a system is shown in Fig. 4 in which annular disks 14, which are of lime-soda glass coated with bright platinum to provide a conducting surface, are mounted in electrical contact with metal bands 16, painted on the inside surface of the glass cylinder. Opposite these bands and on the external surface are similar metal bands 15, which are held at the appropriate potentials by tappings from the potential divider 17. The internal annular electrodes 14 are maintained at their appropriate potentials by electrical leakage between the bands 15 and 16 through the lime-soda glass wall.

A series of ten such tubes of varying length and with varying number of annular electrodes has been consistently free from background. Those having image sections 18 of 8 in. length have been free from background up to approximately 30 kv. The greater the number of baffles the higher the voltage that the tube will stand before some slight background glow appears. Several tubes having an image section 15 in. in length can be operated to 40 kv without visible background. The success of the simple annular baffles in suppressing background is fair justification for the hypothesis on which it was based, but corroborative experiments have not yet been done.

The reliance on conduction through the tube walls to maintain the correct electrode potentials appears to be justified, judging by the observed fact that the electron images formed on the screen are stable, of good geometry, and do not vary in shape or position over wide variations of electron current through the tube.

The vacuum-holding quality of this design of tube is excellent. Hence the photocathode activation is more reliable and when completed the photocathode sensitivity remains constant over long periods. In other respects the performance of the tube, for example as regards light gain and resolution, is very similar to that of the earlier tube.

Thus it can be claimed that a design of tube has been evolved which can be produced and will last reliably. It will give a light gain of between fifty and one hundred times using light of the same (blue) colour at input as at output, an effective output resolution of 30 line-pairs per mm, good geometrical resolution and background illumination which will allow very faint images to be photographed by long exposures. By incorporating a more efficient type of photocathode, for example the tri-alkali cathode, and operating at 40 kv the gain may be considerably improved, especially for white light or light in the wavelength region from 6000 to 8500 Å.

In the U.S.A. and the U.S.S.R. image tubes with mica windows have been made and used for astronomical observations especially in the infra-red region where the image tube is even more competitive with direct photography than in the visible. However, very little has yet been published about the technical details of the tubes.

In the U.S.S.R. Krassovsky and Shcheglov (1958) have reported astronomical observations using electrostatic focusing tubes with curved cathodes and mica windows 10 mm in diameter and 25 μ thick. The resolution achieved is given as 50 μ and the gain over photography is quoted as several orders of magnitude in the infra-red using S-1 photocathodes (Volkov, Yesipov and Shcheglov 1959), 50 to 100 times at H_α wavelength using S-10 or S-20 photocathodes (Shcheglov 1960) and 20 to 30 times over unaided photography in the blue using S-11 photocathodes (Yesipov 1960).

In the U.S.A. Ford *et al.* (1961) have carried out tests and made astronomical observations with tubes made by I.T.T. Laboratories for the Carnegie Tube Committee. The first tube used had S-1 type photocathodes, used electrostatic focusing and had windows 14 mm diameter and 12 μ thick. These gave 30 line-pairs per mm resolution over 5 mm diameter and the gain over direct infra-red photography at about 1·0 μ wavelength was approximately 500 times. Here the resolution compares well with infra-red sensitive plates, but in the blue, similar tubes with an S-11 cathode have a gain of only approximately 10 times over 103a–O plates, partly because of the lack of resolution. This group is now testing magnetically focused tubes with long, narrow, slot-shaped mica windows.

2.4. *Electron-transmitting Mica-window Image Tubes*

The success of the image tube design described in the last section raised the interesting possibility that it might be practicable to operate the tube at sufficiently high voltage and to use a mica window sufficiently thin to allow the photoelectrons to be projected right through it with sufficient residual velocity to be recorded on a photographic emulsion in contact with it. This work has been carried on by Wheeler and the author (McGee and Wheeler 1961) with very promising results.

The first experimental tube was similar to that shown in Fig. 4 but the image section was lengthened to 16 in. with a corresponding increase in the number of annular baffles, and the mica window was 1 cm diameter and 10 μ thick. The inner surface of the mica window was coated with an evaporated layer of aluminium

about 500 Å thick to stabilize its potential. On test this tube could be operated to 45 kv overall potential, without electrical breakdown. When a phosphor screen was painted on the outside of the mica window and an image focused on the photocathode no fluorescent image was visible until the voltage reached about 30 kv. As the voltage was raised above this value the screen brightness increased very rapidly until it was quite bright at 40 kv and a well-defined replica of the image on the photocathode became visible. Thus it appears that electrons of 30 kev energy begin completely to penetrate the mica window and that with electron energies of 40 kev a large proportion of the electrons are penetrating it with considerable residual energy.

One very attractive feature of this type of image tube is that the electron image may be recorded on very fine-grained electron-sensitive emulsion, such as is used for the recording of nuclear particle tracks, for example Ilford type K.2 emulsion. In fact, in a very preliminary test using this emulsion a test image has been recorded showing resolution of approximately 30 line-pairs per mm. With careful adjustment it seems probable that considerably better resolution than this may be attained.

In order to facilitate the adjustment of the tube focus in operation it is proposed to evaporate a thin layer of transparent phosphor on to a small area of the inner surface of the mica window. Observation of a test image on this area will enable the exact focus of the optical and the electron images to be adjusted.

Tests have shown that mica windows of 1 in. diameter approximately 7μ thick will stand 5 to 6 atmospheres pressure. Hence it appears that windows of this diameter and thickness should be safe to use. Efforts will also be made to reduce this thickness further and to increase the operating voltage, since it is essential that as nearly as possible all photoelectrons should penetrate the mica and have sufficient residual energy to produce at least one developable grain of silver halide in the photographic emulsion. Probably the majority of applications for such a tube would be in the recording of faint spectra. For this purpose a slot-shaped window, e.g. 40 mm × 3 mm, would be suitable and this would allow still thinner mica to be used, say 5μ thick, which would give both better resolution and electron transmission efficiency.

If it becomes practicable to record in this way a large proportion of the photoelectrons liberated from the photocathode as a minute clump of fine grains on such an emulsion, it follows that it will be possible to record a very much larger amount of information from a given optical image per unit area of emulsion than has been possible with fast, light-sensitive photographic emulsions. This feature of this method would allow much more information per unit area to be recorded from a given optical image which would give a better signal/noise ratio in the recorded image. This would be particularly useful in detecting very faint objects against the sky background which is the factor which limits the range of very large telescopes in photographing such objects. This method shares this advantage with the image recording tube developed by Lallemand and described below.

2.5. *Lallemand's Electronic Camera*

Photoelectronic image tubes were made experimentally in the early 1930's and were put to use as component parts of television pick-up tubes (Lubszynski and Rodda 1934, McGee and Lubszynski 1939) soon afterwards. It was Lallemand,

however, who seems to have been the first to envisage the possible advantages of using such devices to increase the efficiency of optical image detection. He reported his first experiments in 1936 (Lallemand 1936) and, with his colleagues in Paris, he has continued to improve and use his apparatus from then until today (Lallemand and Duchesne 1951, 1952, 1954).

The basic principle of Lallemand's method is that the photoelectrons from the photocathode are accelerated and focused on to a photographic emulsion which is exposed in the same vacuum chamber. This method was considered impracticable by most vacuum physicists who were familiar with the nature of photoelectric cells,

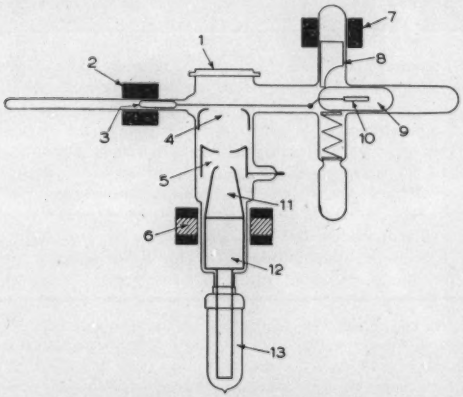


Fig. 5. Diagram of Lallemand's electronic camera.

since it is well known that a good vacuum is essential for the processing and satisfactory life of a photocell. Any organic material within a photocell envelope is fatal to its activation and/or continuing efficiency. The probable reason is that the gases and vapours, especially water vapour, evolved from a photographic emulsion or similar material, react readily with that part of the caesium used to activate the photocathode which remains in the free metallic state on the surface. The conversion of this caesium layer to a caesium compound destroys the photoelectric efficiency of the surface.

The history of Lallemand's work since 1936 has been mainly the record of his struggle to overcome this main difficulty; until today it is fair to say it has been overcome although the method remains experimentally difficult. To him must go the credit for having shown experimentally the large factor of gain in efficiency inherent in the photoelectronic method as compared with direct photography.

Lallemand's apparatus as now used is illustrated in Fig. 5. It consists of a glass envelope which has a removable glass window 1 which can be closed vacuum tight with a suitable O-ring seal. This vacuum envelope is connected to a vacuum pump via a large stop-cock and vacuum joint (not shown) so that it can be detached

from the vacuum pumps. The photocathode 10 which, in operation, will rest on the electrode 4 immediately beneath the window 1 is previously prepared in a separate high vacuum system from which it is sealed off enclosed in a thin glass shell 9. This is introduced into the apparatus through the window 1 and attached by a rod to a magnetic iron slug 3 which can be moved, and so also the photocathode, by means of an external electromagnet 2. After sealing and evacuation of the system the glass capsule is moved to approximately the position shown in the diagram where it rests on an ' anvil ' and can be struck sharp blows by a hammer 8 operated by an external electromagnet 7. In this way the glass envelope is shattered leaving the photocathode exposed. It is then retracted to its operating position, sitting on electrode 4, where the optical image can be focused on to it through the window 1.

Below the photocathode the electrodes 4, 5 and 11 form an electrostatic electron-lens to which an overall potential of about 20 kv can be applied to accelerate the photoelectrons, liberated from the photocathode by the optical image, and focus them to form an electron image on a photographic plate in the magazine 12. This chamber contains eight photographic plates which can be brought in succession into the plane of the electron image, again by the operation of external electromagnets 6 on an internal armature. Thus, eight exposures can be made before the vacuum must be broken and the preparation of the apparatus begun again with new plates and a new photocathode.

The metal case of the photographic magazine has attached to it a metal rod which passes through the glass walls of the tube into a Dewar flask where it can be cooled to liquid air temperature and hence, by conduction, the photographic plates are maintained at a very low temperature. This greatly reduces the evolution of gas and vapours, especially water vapour, from the photographic emulsions which, of course, cannot be subjected to a normal degassing bake. A similar cooling system, not shown in Fig. 5, keeps the photocathode cool during operation. This inhibits the reaction of the caesium of the photocathode with residual gas and also reduces the thermal electron emission by a large factor compared with that at room temperature.

In preparing the tube for use the photocathode ampoule and the photographic plates are introduced and getters are provided in the upper part of the tube. The window 1 is then sealed in place with a silicone O-ring. After preliminary evacuation with conventional pumps the central part of the tube is baked to 400°c during which time the plates and the photocathode are kept cool with liquid air in the Dewar flasks. When evacuation is complete the tap in the tube connection to the pumps is turned off and the apparatus is disconnected from the pumps. The getters are evaporated to remove residual gas, the protecting shell of the photocathode is removed, and the cathode placed in its operating position.

The apparatus is now mobile and can be attached to a telescope, for example. The Dewar flasks are so designed that they can be tilted to a considerable angle without the liquid air flowing out. When prepared and used in this way the photocathode suffers negligible deterioration in photosensitivity in twelve hours, and hence the apparatus is effective for at least one night's work on a telescope.

Figure 6 (Plate I) shows a photograph of the Lallemand electronic camera attached to the Coudé spectrograph of the 120 in. reflector telescope at the Lick

PLATE I

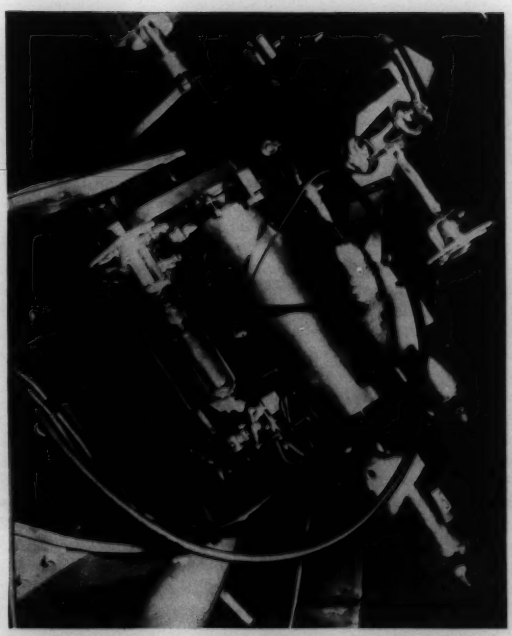


Fig. 6. Lallemand's electronic camera attached to Coudé spectrograph of 120 in. Lick telescope (*by courtesy of* Lallemand, Duchesne and Walker 1960).

observatory where it has been employed successfully by Lallemand, Duchesne and Walker (1960) for several astronomical observing problems. The two Dewar flasks for cooling photocathode and photographic plates are obvious and the main image tube is surrounded by a magnetic screening cylinder to avoid image deterioration due to changing magnetic fields.

The performance achieved can be summarized as follows :

(i) *Resolution.* This is very good, reaching 70 line-pairs per mm in the centre of the field. However, because of the electrostatic lens system used, this falls off towards the edges of the image. This can be compensated for by curving the photocathode plate, concave towards the electron lens, and good resolution can be preserved over three-quarters of the diameter of the field. The curvature of the photocathode plate must be allowed for by using a correcting field-lens in the optical system. This is the main disadvantage of the electrostatic electron-lens system. It is claimed that the resolution is limited by the grain of the photographic plate although it is the very fine-grained nuclear emulsion (for example Ilford type G.5) that is used. Emulsion of this type is the most efficient for recording electrons which, if they have energies of about 20 kv, will each produce a cluster of several developable silver halide grains. Thus, each electron is recorded with great certainty and as there are very many more grains per unit area (about 100 times) in this emulsion than in normal fast, light-sensitive film, it follows that many more light photons of an image can be recorded linearly per unit area on the former than on the latter. This results in a better ' signal/noise ratio ', which appears in practice as an improved discrimination of faint objects against a bright background.

(ii) *Spectral range covered.* Using an antimony–caesium photocathode this is from the ultra-violet cut-off of the glass window, $\lambda \sim 3500$ Å, to the long-wave threshold of the cathode at $\lambda \sim 6700$ Å. The maximum response is between 4500 and 5000 Å. This range is considerably greater than the range covered by the best, fast, photographic plates, for example Eastman Kodak 103a–O.

(iii) *Linearity of record.* Because the response to light of the photocathode is quite linear and, over a very large range of recorded information, the ' exposure ' of the fine-grained plate to electrons is also linear, the overall density recorded is linear with respect to the light of the image.

(iv) *Sensitivity gain.* When using the nuclear emulsion and using blue light the gain in sensitivity is between 50 and 100 times compared with direct photography using an Eastman Kodak 103a–O emulsion. The variation of gain is, no doubt, largely dependent upon the sensitivity of the particular photocathode used. Again the resulting electronic camera picture does not suffer from granularity of the emulsion to the same extent as the conventional light-sensitive plate.

Lallemand and his colleagues have recently shown (Lallemand *et al.* 1960) that if the number of background electrons is kept small by cooling the photocathode and the number of photoelectrons is kept so small that the individual tracks do not overlap, then by a process of counting the individual electron tracks, a gain in sensitivity by a factor of between 10^3 and 10^4 compared with direct photography can be achieved.

These data are very impressive and it is only the practical difficulties involved, especially in the earlier forms of the apparatus, that has deterred people, especially astronomers, from using it. However, it has been used very successfully by

Lallemand's French astronomer colleagues, for example Wlérick, Fehrenbach and Rösch. It remains to be seen whether or not some device can be developed which will give results comparable with those obtained with Lallemand's apparatus while being simpler to use.

2.6. *The Two-chamber, Single-stage Image-recording Tube*

This device is in principle the same as Lallemand's apparatus but employs a different method of preserving the photocathode sensitivity. Tubes of this type were first investigated by Hiltner (Miller, Hiltner and Burns 1956, Hiltner 1956) and his colleagues of Yerkes observatory and by Baum (Hall, Ford and Baum 1960) and his colleagues of the Carnegie Image Tube Committee, the tubes being made by various American industrial laboratories.

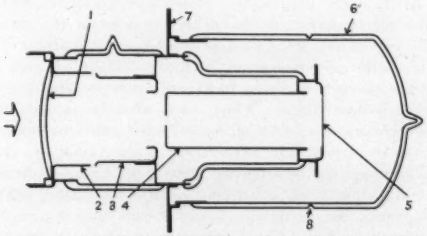


Fig. 7. Two-chamber, single-stage image recorder (*after* Hall, Ford and Baum 1960).

The principle of this method is to divide the vacuum chamber into two compartments by means of a thin, vacuum-tight film. In one compartment, in which the photocathode lives, very good vacuum is maintained. The vacuum in the other compartment is only moderately good but the photographic plates can be introduced into it through a vacuum lock, exposed to the electron image, and then removed again. The photoelectrons are accelerated, projected through the thin film and form an image on the photographic emulsion which is usually brought into close proximity to the thin dividing diaphragm.

A representative tube of this type is shown in Fig. 7 in which the single solid lines represent metal walls and the double lines represent glass parts of the tube. The photocathode 1 is deposited on a glass window curved slightly convex to the optical system. This is necessary since an electrostatic electron-lens system is used, formed by the electrodes 2, 3 and 4, in order to achieve a reasonably flat electron-image field.

The thin film 5, about 800 Å thick, of aluminium oxide or aluminium metal stretched across the aperture is intended to form a vacuum-tight, though very fragile, barrier between the two compartments of the tube. The right-hand section of the tube envelope 6 is sealed to a metal flange 7 which can be clamped in a vacuum, plate-changing chamber which covers the whole of the right-hand section of the tube. After the tube has been processed and fixed in the plate changer, and the latter evacuated, the glass shell 6 is removed by cracking the glass cap round the groove 8 by a suitable heat shock from a hot filament. The

photographic plates can then be brought up into close proximity to the thin dividing membrane. Because of the fragility of the film it has not yet been found practicable to bring the plate into actual contact with the film. The separation between plate and membrane was approximately 40 μ. Because of the scattering of the electrons in the film, this results in a maximum recorded resolution of about 10 line-pairs per mm.

Results obtained by Hiltner (1956) and his co-workers from tests with a very similar type of tube gave a limiting resolution of 15 line-pairs per mm, again with the photographic emulsion spaced 500 μ from the dividing diaphragm. The gain in recording speed as compared with direct photography was about fifty times, although the photocathode of the tube was at the time rather insensitive, about 15 μA/lumen.

The tube was operated at 15 kv and at this voltage the exposure had to be restricted to a few seconds, otherwise the plate was blackened appreciably by the spurious background. This trouble is probably aggravated by the presence of caesium in the working part of the tube, no precaution having been taken to exclude it.

There is a tendency for small air leaks to occur through the thin dividing diaphragm in this tube. Hence the vacuum in the photographic plate chamber must be maintained at a very low level, otherwise the air leak through the diaphragm will destroy the sensitivity of the photocathode. This is a serious inconvenience in the use of the tube since a vacuum pumping system must operate continuously once the tube has been put into operation.

2.7. *Electrostatic Focusing, Single-stage Tube*

This general design of image tube is illustrated in Fig. 8. It was developed during the war for infra-red image detection by the Radio Corporation of America (Morton and Flory 1946) and in Germany by the A.E.G. (Eckart 1937), and they are surprisingly similar in basic design. Fig. 8 shows diagrammatically the design of the R.C.A. tube.

The photocathode is formed on the spherically curved surface 1 of the glass end of the tube. Electrodes 2, 3, 4 and 5 form an electrostatic immersion electron-lens which focuses the photoelectrons liberated by light from the photocathode 1 onto a flat phosphor screen 6 which is held at about 4 kv positive relative to the photocathode. The end of the tube 7 is a hemispherical glass lens which in combination with a second plano-convex lens forms a simple magnifying eyepiece of about 10 times power. The photocathode surface 1 is curved convex towards the electron lens in order to achieve a reasonably flat electron-image field in the plane of the phosphor screen. Without this the non-uniformity of image resolution and geometrical distortion would be unacceptable. Hence the incident optical image-plane must be curved to fit it to the curved photocathode. With the best compromise, the resolution in the centre of the image is about 250 line-pairs over the whole screen and at the edge it falls to about 150 line-pairs defined in the same way. There is considerable pin-cushion geometrical distortion.

Important features of the operation of such a tube are that it does not require a stabilized current for magnetic focusing and the high voltage supply need not be very carefully stabilized since the focus of an electrostatic electron-lens depends

mainly on the ratio of the potentials applied to the component electrodes, and these are determined by a potential divider. It is only affected to a second order by changes in the overall potential.

This design also lends itself very conveniently to taking advantage of the following useful feature of electron-image intensifiers (Brüche and Busch 1937, McGee and Lubszynski 1940). If the dimensions of the electron image are reduced between the photocathode and the phosphor screen, the resulting image is increased in brightness for two reasons : first, the total amount of light liberated by the electrons is greater than that required to produce them and second, this light is liberated from an image of smaller dimensions. Hence the light can be used more efficiently

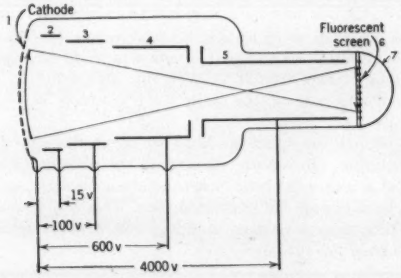


Fig. 8. Single-stage electrostatic focusing image intensifier (*after* Morton and Flory 1946).

for observation, either by the eye or by photographic recording by a conventional transfer optical system. Thus if the genuine light gain is $\times 50$ and the linear image demagnification is $\times 5$, the image brightness gain is $\times 1250$. However, spurious background appears seriously to reduce the effective gain that can be achieved since above about 15 kv operating voltage it reduces the image contrast. This basic electrode design is the same as that used in the two-chamber tube (§ 2.6) and Lallemand's electronic camera (§ 2.5). In the former of these, background is a serious limiting factor but in the latter, where free caesium does not contaminate the electrodes, this type of background appears to be absent.

This same general design is used in several other image intensifiers : in particular it is the basic design of each stage in multi-stage intensifiers which will be described later (§ 3.2) and in the x-ray image intensifier which will now be described very briefly.

2.8. *X-ray Image Intensifier*

(See Teeves and Tol 1952.)

This is a tube with substantially the same electrode design as that shown in Fig. 8. The main novel feature is the cathode which is a spherically curved transparent glass screen about 5 in. in diameter, carrying a layer of x-ray sensitive phosphor on one side, backed by an aluminium film, and a transparent photocathode on the other side—the side facing the electron-lens system. The glass

envelope at the cathode end is usually a hemisphere of glass made very thin to minimize x-ray absorption. The x-rays pass through the body to be examined, then through the thin window to form a shadowgraph on the phosphor screen. The fluorescent light from the x-ray image on the phosphor screen passes through the photocathode from which it releases electrons, which are accelerated and focused on to the phosphor screen.

If the light gain in the one stage of image intensification is $\times 50$ and the image brightness is increased by $\times 25$ by a linear reduction in dimensions of 5, the overall gain should be about 1000 times. This appears not to be achieved, in the sense that the incident x-ray intensity can be reduced by this factor. In practice it appears that the x-ray intensity can be reduced by a factor of about 10. The failure to achieve the theoretical factor can probably be accounted for, partly by short-fall in photocathode and phosphor efficiencies and, more important, by loss of image contrast due to spurious background arising at the high voltage, of about 25 kv, at which the tube must be operated to achieve adequate gain.

This type of tube has been of great value in x-ray radiology, enabling the radiologist to reduce the radiation dosage considerably. Its main defects are the small effective area of the x-ray sensitive screen and the fact that the radiologist must observe through an eyepiece lens. Naturally still greater efficiency would be desirable, covering a bigger useful field and with a television screen presentation.

2.9. *High-speed Shutter Image Tube*

One of the most important observational functions of image tubes is to act as high-speed shutters or high-speed display tubes. In the first case the tube may be a simple diode in which the whole potential difference between cathode and anode is pulsed on and off in short time intervals. If, however, adequate light gain is required from such a tube this potential difference must be large, not less than 10 kv, and it is not easy to change the voltage of electrodes, which have considerable capacity to adjacent electrodes, through a large range in very short time intervals, for example about 10^{-7} sec. An improvement can be made by constructing the tube as a triode with an electrode intermediate to the anode and cathode which will operate as an electron gate with a change in potential of a fraction of that between cathode and anode. Thus a fine mesh electrode might be stretched in front of the cathode and at a distance from it of, say, 1 mm. However, with a field gradient of 1 kv cm^{-1} this would require a potential swing of about 100 v to stop all photoelectrons. Moreover, the mesh would have a considerable capacity to the cathode and this would demand driving circuits of high power to achieve high-speed gating.

A design of image tube for achieving very high-speed gating recently developed in the author's laboratory by Mandel (1961) is as follows. The tube consists of two sections. In the first, the photoelectrons are accelerated from the photocathode for a short distance to an energy of about 100 ev and they then enter a retarding field and are slowed down until they reach a mesh of very fine pitch and low shadow ratio, held at approximately the potential of the photocathode. The electrons are controlled by an axial magnetic field which constrains them to travel in cycloidal paths about the magnetic lines of force. The extension of this uniform magnetic field brings the electrons to a focus on the fluorescent screen at the

other end of the tube. If the potential of this mesh is only a few volts negative relative to that of the cathode (allowing for contact potentials) all the photoelectrons will be stopped by the mesh and will return to the photocathode. On the other hand, if it is only a few volts positive, almost all the photoelectrons will pass through it into the second section of the tube where they are accelerated and focused onto the screen of the tube. Thus, the voltage swing of the control electrode need not be greater than about 5 volts and its capacity to neighbouring electrodes will be small. Hence it is comparatively easy to apply the potential changes necessary to stop, or let through, the electron stream in very short time intervals.

The voltage swing required on this control mesh is determined to a large extent by the field gradient in the space following the mesh. The larger this gradient, the more it penetrates through the mesh apertures and hence the greater the negative potential necessary to stop the passage of electrons. The field in the last section of the tube is usually quite high, approximately 1000 v cm^{-1}, and this tends to give too strong a field penetration. Hence a second mesh is used following the control mesh to reduce this field to a suitable value of about 100 v cm^{-1}.

A fine metal mesh† of 200 meshes per inch and low shadow ratio, not greater than 15%, is used as the control electrode. It is found that this mesh stops a proportion of the electrons considerably smaller than that corresponding to its shadow ratio. This is thought to be due to the penetration of the field of the second, high voltage, section of the tube through the mesh so as to direct electrons through the apertures that otherwise would have impinged on mesh wires. While this effect improves the electron transmission efficiency of the system it also has the effect of impairing the final image focus because of the lens action of the individual mesh apertures. Control electrodes of this type are incorporated into the tube shown in Fig. 15 (Plate III) as electrodes 2 and 3.

§ 3. Multi-stage Image Intensifiers

3.1. *General Considerations*

In the single-stage image-intensifier tubes described in § 2, the gain was achieved by increasing the energy of individual electrons. Those methods will now be discussed in which the individual image photoelectrons are multiplied in number by a large factor and then the 'multiplied' electron image is accelerated and recorded. This process is illustrated in Fig. 9. The tube consists of a glass cylinder 1 about 2 in. in diameter, having glass windows 2 and 3, at either end. On the inner surface of the window 2 is a transparent conducting photocathode 4 on to which the optical image is focused by a lens 15. The liberated photoelectrons are accelerated by a suitable axial electrostatic field produced by suitable potentials from the potential divider 14 applied to electrodes 5 mounted on the tube walls, and focused by an axial magnetic field produced by a current flowing in the solenoid 6.

The focusing process is identical with that for the single-stage tube described in § 2.1. If the electrostatic accelerating field and the magnetic field are accurately parallel to the axis and uniform, then slow photoelectrons leaving the cathode 4

† Kindly supplied by Messrs. E.M.I. Electronics Ltd.

can be accelerated and, by suitable adjustment of the two fields, focused in the plane of the first intensifying screen 7. The quality of the electron image resolution will be uniform over the whole screen 7 and image geometry will be exactly reproduced. However, in practice, because we cannot make the coil infinitely long and the electrostatic field must be applied in steps, there is some loss of definition towards the edge and distortion appears in the image which results in diagonals through the centre of the image being distorted into an S-shape, commonly known as S-distortion.

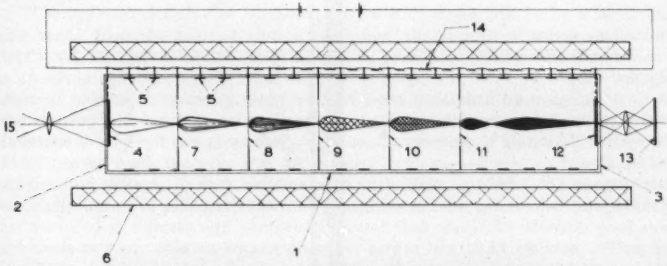


Fig. 9. Diagram of multi-stage electron multiplier image intensifier.

If each photoelectron that reaches the screen 7 can produce several slow electrons from the opposite side of this screen, these can be accelerated to, and focused on, the next screen 8, and so on through screens 9, 10 and 11. If an average multiplication factor of m is achieved at each stage and there are n stages, then m^n electrons will leave screen 11. These are accelerated to high energy and focused on a fluorescent screen 12 to produce a flash of fluorescent light which is photographed on the plate 13.

If the multiplication factor m^n is large enough and the energy of the electrons when they reach the screen high enough, then the descendants of a single photoelectron can produce enough photons from the output phosphor to enable it to be recorded as a spot on the photographic plate. The two principal methods for achieving electron multiplication at the screens 7 to 11 will now be outlined.

3.2. *Cascaded Phosphor–Photocathode Intensifying Screen Tubes*

This method was first suggested in a patent originating in the Philips Laboratories in 1928 (Philips Gloeilampenfabrieken 1928). The structure of this type of screen is illustrated in Fig. 10. It consists of four layers : first, a thin transparent sheet A of glass or mica, about 5 μ thick ; second, on the side on which the primary electrons impinge, a layer B of phosphor which also must be about 5 μ thick ; third, an almost opaque film of aluminium D about 800 Å thick, deposited on the phosphor to reflect all fluorescent light through the transparent support plate ; and fourth, the transparent photoelectric cathode C which should have a response to light that matches the colour of the light output from the phosphor. Probably

the most efficient combination is ZnS:Ag, or P-11, phosphor with antimony–caesium, or S-11, photocathode as their emission and response curves respectively have maxima at about 4500 Å and both are very efficient. However, it is possible that the tri-alkali [Sb]–K–Na–Cs, or S-20, cathode combined with a ZnS:CdS phosphor might give a more efficient intensifier combination.

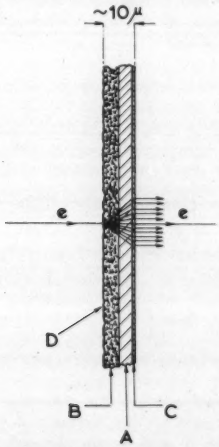


Fig. 10. Cascaded phosphor–photocathode intensifier screen.

Since the second-generation electrons produced at such a screen are relatively slow ($\gtrsim 1$ ev) photoelectrons, it follows that they can be brought to a focus on the next screen 8, in Fig. 9, with an adequately small circle of confusion, the diameter of which varies directly with the distance and inversely with the voltage between the screens 7 and 8.

Both the phosphor layer and the transparent supporting sheet must have thicknesses that are not negligible compared with the image definition required. The phosphor must be about 5 μ thick in order to absorb substantially all the energy of the incident electrons, which may be 10 to 25 kev. The supporting sheet will probably need to be at least 5 μ thick to have sufficient mechanical strength to survive the deposition of a phosphor screen and the aluminium backing. The photocathode layer thickness is quite negligible by comparison, probably of the order of 500 Å thick. Hence the light emitted isotropically by the phosphor due to the impact of one electron will spread over an area of phosphor about 5 μ in diameter; it will spread to about 10 μ diameter in passing through the transparent sheet to the photocathode and hence the daughter electrons due to this one primary electron will originate from within a circle of about 10 μ diameter. A further loss in resolution will result from the aberration in the electron-optical focusing caused

by the finite initial velocities of the photoelectrons. From experiments on single-stage tubes it appears that this effect is likely to be small since under good conditions resolution of 100 line-pairs per mm has been achieved.

The results from the thin window image tubes (§ 2.2) give a guide as to what may be expected. In those tubes the mica window was about 12 μ thick, the image resolution on the phosphor was 60 line-pairs per mm and 30 line-pairs per mm on the contact photograph. Hence it seems that a reasonable estimate of the transferred resolution through a dividing sheet 5 μ thick would be between 40 and 50 line-pairs per mm.

The expected decrease in resolution after n such stages will be by a factor $\sqrt{n}$ and hence, from the point of view of resolution, the number of stages should be kept small and the voltage per stage as high as possible. On the other hand there is an advantage to be gained from using more stages for the same overall voltage since as soon as the primary electrons have been multiplied by a useful factor n the electric field can be used more effectively because there are n times as many electrons to which it can give energy. Thus, for example, if an overall voltage of 40 kv can be used, the tube may be divided into two sections, each with 20 kv accelerating voltage, or four sections each with 10 kv applied. We assume the same efficiencies of photocathodes and phosphors in both cases: that ten photons are required to liberate one photoelectron and that one 20 kv electron produces 2·5 times as many photons from a phosphor screen as a 10 kv electron. That is, if a 20 kv electron produces 25 electrons from an intensifier screen, a 10 kv electron will produce 10. Hence the electron gain in the two-stage tube would be approximately 25 times, while in the four-stage tube it would be approximately 1000 times. Then in the final screen the fast and the slow electrons would produce about 1250 and about 500 photons each respectively. Thus the overall gain would be approximately 50×10^4 for the four-stage and approximately 3×10^4 for the two-stage tube. However, if we assume that the diameter of the circle of confusion of the electron image in each stage is D the final circle of confusion on the last screen will be $\sqrt{2}\,D$ in diameter in the case of the two-stage tube and $\sqrt{4}\,D$ in the case of the four-stage tube ; that is, the *area* of the spot resulting from one primary electron in the four-stage will be twice that obtained in the two-stage tube. The bigger diameter of the spot the less efficient the recording on a photographic emulsion since, with the same transfer optics, the image spot will cover twice the area. Hence this factor of 2 is in favour of the two-stage tube.

As noted above the circle of confusion for each electron scintillation is likely to be $\sqrt{2}$ times greater in the four-stage than in the two-stage tube and hence, for a tube of the same effective diameter, it will reproduce twice the image detail.

A further advantage to be expected from using fewer high voltage stages compared with more low voltage stages is that the multiplication factors achieved by individual electrons will be larger in each stage in the former case and hence the mean percentage deviation will also be smaller. If the average multiplication at the first screen is 25 in the high voltage and 10 in the low voltage stage, the mean deviation will be $\sqrt{25}$ or about 20% in the former and $\sqrt{10}$ or 30% in the latter. This will be repeated in subsequent stages and the relatively greater spread in multiplication factors for the low voltage stages will be still further increased.

Moreover there are more stages in which to continue the spread. Now it is important that all pulses of electrons that reach the phosphor screen 12 in Fig. 9 should be as nearly as possible equal, so that the records produced by each should be of nearly equal weight. Otherwise a further source of noise is introduced which reduces the effective signal/noise ratio below that ideal maximum value determined by the shot noise inherent in the primary photoelectron stream. Hence, it follows that the arrangement of few, high voltage stages is to be preferred from this point of view though the actual weighting factor to be attributed to this effect is difficult to estimate.

If we allow a factor of advantage of 2 on this account and take into account the factor of 2 resulting from loss of definition discussed above, it is clear that the advantage of the multi-stage tube over the few-stage tube is not as marked in respect of gain and it loses on the score of definition. Hence it seems that the choice between these two designs will be settled mainly by the degree of perfection with which they can be put into practice. The multi-stage tube requires more components, for example photocathodes, screens, etc., to be made with reasonable perfection inside one envelope than the two-stage, but the latter must stand up to the application of quite high voltages, about 20 kv per section, without appreciable background resulting.

Another feature of this type of intensifier tube is the fact that on each of the intensifying screens the fluorescent image persists for a period of time determined by the decay rate of the phosphor. This provides a useful, short-term memory which can be employed under the following circumstances. Suppose it is necessary to record an event, such as a lightning flash of short duration, which happens at an unpredictable instant. If the image tube is kept in operation observing the place where the event is to occur continuously the recording film is likely to be over-exposed by build-up of background alone before the wanted event takes place. This can be overcome by maintaining the first stage only of the image tube in operation continuously, so that a short-lived image is maintained on the first phosphor screen, but the subsequent stages are inoperative. Hence no light will reach the film. Then, when the unpredictable event occurs, a signal can be generated, e.g., from a photomultiplier, enabling all stages of the image tube to be energized so rapidly that before the evanescent fluorescent image on the first intensifying screen has decayed appreciably it is relayed through the rest of the stages to the final screen and photographed.

Image intensifier tubes of this general type have been made and used by several groups. The first really effective tube appears to be that developed and described by Zavoiskii *et al.* (1955, 1956). This was a six-stage tube with electrostatic electron optics in the first stage and means for suppressing the electron stream or scanning the electron image over the first intensifying screen. The following five stages were electromagnetically focused. Little detail was published about the tube but it was capable of achieving sufficient light gain to allow the flashes produced by individual photoelectrons to be photographed. The accelerating voltage across the whole tube appears to have been pulsed on for short instants to record unpredictable, evanescent light images. In this way they were able to record pictures of the tracks of nuclear particles passing through transparent blocks of phosphor; faint, minimum-ionizing-particle tracks appearing as a row of separate, very similar dots, each corresponding to a single photoelectron.

Later Butslov *et al.* (1958) used the same type of tube for astronomical tests. In these experiments a multi-stage tube with electromagnetic focusing throughout was used and it was operated at 6 kv per stage to reduce background. Nevertheless, the gain was sufficient to allow single photoelectrons to be recorded and the tube background was negligible compared with that due to the sky. By comparison of direct photographs of stars with photographs obtained by means of this image intensifier the authors claim a gain in speed by a factor of 1000 times, but the photographs on which this claim is based are not published. The pictures they publish do appear to suffer from ' graininess ' and lack of image definition.

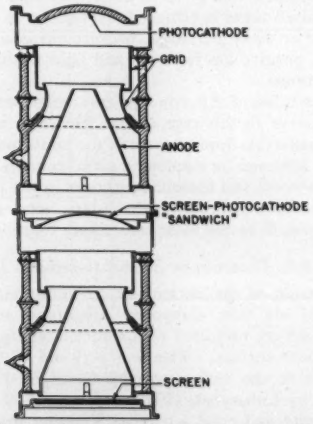


Fig. 11. Two-stage, electrostatic focusing image intensifier (Radio Corporation of America, Stoudenheimer 1960).

Work was carried out on a two-stage type of tube by Davis (1959) in the author's laboratory but though promising results were obtained it was inconclusive. The main object of this work was to devise ways and means of constructing a two-stage tube incorporating similar techniques to that employed in the single-stage tube (§ 2.2) so that caesium could be excluded from the parts of the tube, especially the first stage, where it would tend to cause background at high operating voltages. This project is being continued.

Cascade image tubes of two and three stages have been developed by Stoudenheimer (1960) of the Radio Corporation of America using electrostatic electron-image focusing. A two-stage tube of this type is illustrated in Fig. 11. As in the single-stage tube with electrostatic electron optics, the photocathode surface is convex towards the object lens, and the optical image must be corrected to be in focus over this curved surface by means of a suitable field lens. This can be done fairly successfully, but the correction of the electron image is not very satisfactory. In a plane image, definition falls off appreciably from the centre

towards the edge and there is considerable pin-cushion distortion. These defects are aggravated in the two-stage tube since in this the first intensifier screen must be curved in the same sense as the primary photocathode in order to correct the electron image on the output screen. Hence it must be curved in such a way as to further impair both resolution and geometry of the electron image formed on it. Thus the resulting image impairment is rather greater than what would be expected from the product of that due to each section alone.

The image resolution quoted for such a tube is 18 line-pairs per mm at the centre, falling to 4 line-pairs per mm 10 mm from the centre of the image. The image geometry distortion is also given as 20% at 10 mm from the centre. A light gain of several hundred times is achieved, depending on the voltage per stage. However, above 10 or 12 kv per stage, background is said to increase rapidly due to field emission, positive ion feedback and light feedback from glass bombardment and external corona.

Three-stage tubes of this same general design were planned but in view of the difficulties inherent in this type it does not seem very profitable. It should be noted that considerable improvement in the performance of this type of multi-stage tube might be achieved by employing plano-concave fibre-optics screens to couple the phosphor screens and the photocathodes in the intensifier screens. However, for most purposes the uniform electrostatic accelerating and magnetic focusing fields would seem to be the most satisfactory solution.

3.3. *Transmission Secondary-emission Image Intensifier*

In this version of the electron-image multiplying tube shown in Fig. 9 the screens 7 to 11 are thin, composite membranes into which, on one side, the primary electrons are projected with sufficient energy to eject secondary electrons from the opposite surface. This is a very old idea. The first proposal of this method known to the author was by Orvin in 1934. Later it was proposed independently by Lubszynski (1935) and the use of composite films to give both good conductivity and good secondary emitting properties was proposed by the author (McGee 1937). More recently, more detailed structure of the necessary composite films was proposed by Sternglass (1955) and multi-stage image intensifiers incorporating this last proposal have been described by Wachtel, Doughty and Anderson (1960). Work along these same lines has been done in the Instrument Technology Department of Imperial College by Wilcock, Emberson and Weekley (1960 a, b) which has given very striking results. The following outline is based largely on their work.

The design of the tube is substantially as shown in Fig. 9 while the structure of the films is shown in Fig. 12. First a thin film of aluminium oxide A is made by the conventional method of anodizing aluminium foil and dissolving off the residual metallic aluminium. This film is about 500 Å thick and is mounted on a glass ring. A thin conducting layer of aluminium B, about 200 Å thick, is then evaporated in a vacuum on to one side of this film and makes electrical contact with a metallic coating on the glass supporting ring. A layer of potassium chloride C about 500 Å thick is then evaporated on to the surface of the aluminium layer. The functions of these layers are as follows: the aluminium oxide layer is simply a convenient, thin, tough support medium for the potassium chloride layer which is

the actual secondary emitting layer, while the aluminium layer is a conductor which, since both the aluminium oxide and potassium chloride are highly insulating, is required to maintain the layer at its correct operating potential.

In operation the incident electrons enter the membrane through the aluminium oxide layer, with about 5 kev energy, penetrate through the aluminium oxide and aluminium layers and are brought to rest in the potassium chloride layer. Ideally none of these primary electrons should emerge from the opposite surface, unless with very low velocities. The potassium chloride is a very good secondary electron

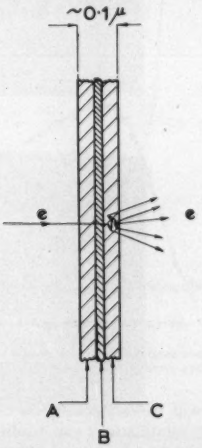


Fig. 12. Structure of transmission secondary emission intensifying screens.

emitter; that is, primary electrons passing through it excite many secondary electrons which have a long mean free path enabling a useful number, for example 5 to 10, to escape from the free surface. Hence it is required that the primary electrons should be brought to rest as near to the free surface of the potassium chloride as possible. At this stage in its path the direction of travel of a primary electron will be almost random, as also will be the directions of movement of the excited secondary electrons. The relations between the energy of the primary electrons, the fraction that penetrate the film and the secondary emission ratio are given in Fig. 13. These curves, for which the author is indebted to Wachtel, Doughty and Anderson (1960) of the Westinghouse Research Laboratories, refer to a layer of potassium chloride on an aluminium film. However, the curves for the film composed of aluminium oxide, aluminium and potassium chloride are believed to be very similar. It is clear that the energy of the primary electrons must be chosen to give as large a multiplication factor as possible consistent with a small percentage of them penetrating the film with appreciable residual energy. Any such fast primary electrons would not be focused in the next stage and would lead

to loss of image contrast. It is found that a stage potential of about 5 kv gives an electron gain of about 5 per stage with only a few per cent of transmitted primaries. Thus, using five multiplying stages, a single photoelectron can produce about 3000 electrons in the last stage. These electrons are then focused in a small spot on the final phosphor screen with an energy of about 15 kev per electron. Each of these electrons will excite about 500 photons from a ZnS:Ag screen and hence

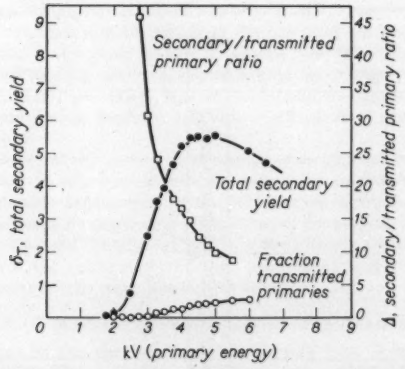


Fig. 13. Relationships between the energies of primary electrons and their penetration of the film and the production of secondary electrons (*after* Wachtel, Doughty and Anderson 1960).

each primary photoelectron will be responsible for the emission from the screen of about 10^6 photons. Such scintillations can readily be seen with the aid of a suitable eye-piece lens and can be photographed with relatively simple transfer optics.

If we assume that the light of the input image is the same quality as that emitted by the ZnS:Ag phosphor screen, it is to be expected that, on an average, one in ten of the incident photons will produce a photoelectron. Hence, an overall light gain of about 10^5 is to be expected. This has been achieved, and those individual photons which happen to produce a photoelectron can be recorded as small spots on a photographic plate. This is well illustrated in Fig. 14 (Plate II), which shows five photographs of the screen of such a tube taken under the following conditions: the exposure time was 0·5 sec and in (*a*) no light fell on the photocathode, while in (*b*), (*c*), (*d*) and (*e*) a fan pattern of black and white lines was projected on to the cathode with gradually increasing light intensity. The white specks in (*a*) represent single thermal, or other dark current, electrons from the cathode. In (*b*) the incidence of a very faint image results in an increase in the density of specks in the area covered by the image, but no detail, and only a rough idea of overall shape can be discerned. In (*c*) the shape of the image and some suggestion that it is made up of lines begins to show. In (*d*) the pattern emerges fairly clearly and in (*e*) the signal/noise ratio of the image is quite good and a good estimate of the limiting

PLATE II

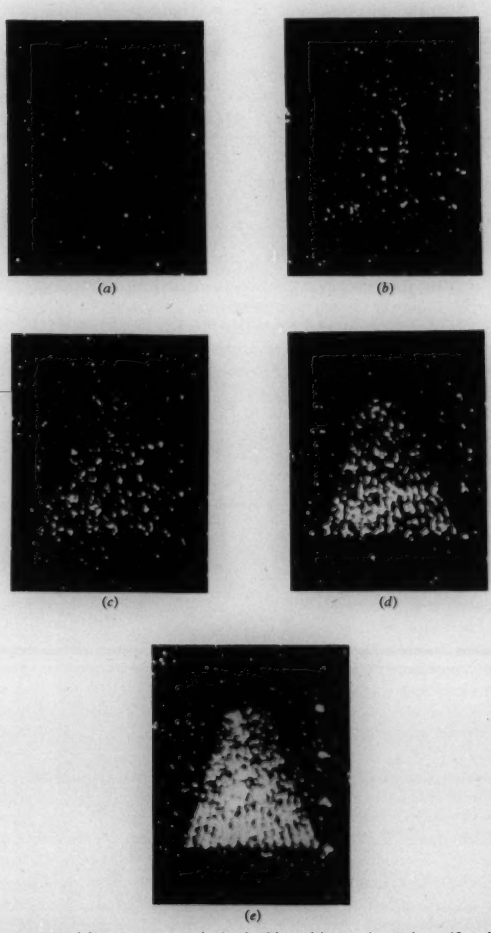


Fig. 14. Photograph of fan test pattern obtained with multi-stage image intensifier showing single-electron scintillations as light increased from zero at (*a*) to maximum at (*e*) (*after* Wilcock, Emberson and Weekley 1960 a, b).

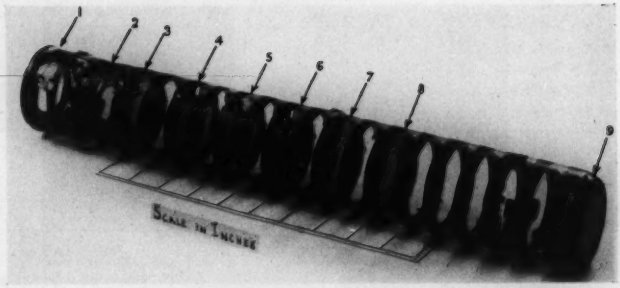


Fig. 15. Photograph of five-stage transmission secondary emission image intensifier with gating electrodes. 1, cathode ; 2, 3, gating electrodes ; 4–8, multiplying films ; 9, phosphor screen.

Fig. 16. Photograph of Westinghouse ' Astracon '.

resolution can be made, which is approximately 15 line-pairs per mm. This corresponds with the observed diameter of individual scintillations of about 40 μ.

In this method of intensification, the thickness of the multiplying screens, which is about 0·1 μ, can make little contribution to the increase in the disk of confusion of each electron bunch at each film. This is primarily due to the considerable spread in velocity and direction of the secondary electrons which are liberated from the film (Kollath 1941, 1947). The distribution in direction is probably approximately Lambertian, while the proportion with energy less than 2 ev is about 50% and falls off rapidly for higher energies. Thus possibly 10% have energies greater than 10 ev. These electrons with such a wide range of initial energies and directions cannot be brought to a precise focus and hence produce a disk of confusion of appreciable diameter.

Addendum : It is possible that space-charge also is a factor limiting the definition since in the last stage about 5×10^3 electrons may leave an area of about 3×10^{-6} cm^2 of the last dynode in about 10^{-10} sec. This is a current density of 3 A cm^{-2} which is certainly high enough to cause space-charge repulsion to be significant.

The overall magnification from photocathode to screen of such a tube is very approximately unity, and in tubes now being made the effective image diameter is about 20 mm. The resolution is fairly uniform across the whole field and hence the image definition corresponds to about 600 picture points across the picture width. In order to increase this it seems necessary to increase the area of the screens, which seems quite feasible, or to shorten the stages, keeping the potential difference the same and so increase the field gradient, to which the disk of confusion due to chromatic aberration is inversely proportional. However, this may increase the background.

As in all image intensifiers, background is of first class importance since it determines the limit of detectability of very faint objects. In the first successful tube of this type, Wilcock and his co-workers measured a background of about 6×10^3 scintillations cm^{-2} sec^{-1}, or 10^{-15} A cm^{-2}, of dark current. But it is known that the genuine dark current from an antimony–caesium photocathode can be at least two orders of magnitude lower than this and hence it was suspected that some of this dark current was spurious. In fact, in later tubes of better design and operated under more favourable conditions the dark current count has been as low as about 100 scintillations cm^{-2} sec^{-1}.

A serious doubt regarding the useful life of such a tube using potassium chloride as the secondary emitting material was raised by the Westinghouse group who found that it suffered rapid fatigue under electron bombardment. Hence they carried out a search for materials which did not show fatigue, and magnesium oxide and barium oxide were found to be better from this point of view. However, life tests by Wilcock and his co-workers have failed to show any serious fatigue effect which would render the useful life of such a tube unduly short.

A photograph of a tube as made by Wilcock is shown in Fig. 15 (Plate III). This shows the photocathode 1, the five multiplying films 4, 5, 6, 7 and 8, and the phosphor 9. In this tube there are also the two mesh control electrodes 2 and 3 for gating the tube as described in § 2.9.

The tube of this type manufactured by the Westinghouse Company under the trade name of ' Astracon ' is shown in Fig. 16 (Plate III). The tube is constructed

of metal flanges which hold the multiplying films, sealed to and spaced by short ceramic cylinders. The films were spaced 0·3 in. apart and the cathode and anode spaces were 0·6 in. Hence, overall, the tube is much shorter than that designed by Wilcock. Since the voltage per stage must be about the same, the magnetic field must be considerably greater. This is produced by a permanent magnet and the shorter length of the tube makes this easier but the electrical field gradient must be steeper and hence the danger of spurious emission must be greater. The definition should be better.

Another feature of the Astracon is that the secondary emitting films are supported on low shadow-ratio grids for mechanical strength. This has the unfortunate result that the grid wires tend to appear as dark lines on the image and to produce moiré patterns.

Problems and possible defects in this device which should be mentioned are :

(i) Some primary photoelectrons will be stopped in the first film without producing even one secondary electron and hence the information corresponding to these will be completely lost. Similar, but progressively smaller losses occur in subsequent films. The extent of this loss is not known at present.

(ii) As the multiplication factor is about 5 at each film, the mean deviation from this will be large—probably not less than $\sqrt{5}$ or 45%. Hence there will be a wide spread in the multiplication factor for individual electrons in the first stage which will be increased, though to a lesser degree, in each succeeding stage. This will result in a wide spread in intensities of the scintillations and a decrease in the effective signal/noise ratio in the final recorded image. This shows clearly in the photographs of Fig. 14. The magnitude of this effect is at present the subject of investigation.

(iii) There is some evidence that the potassium chloride is disintegrated by electron bombardment and the chlorine released in small quantities. If this can reach the photocathode it combines with the caesium and destroys the sensitivity. Fortunately, this evolution of gas will be quite small from the nearest, and a maximum from the most remote, film. However, precautions should be taken in this design to prevent the diffusion of chlorine back to the cathode.

(iv) The emission of secondary electrons is almost instantaneous, unlike the emission of fluorescent light, and hence there is no effective memory mechanism in these tubes except the phosphor screen. Hence it is necessary to use two tubes of this type in order selectively to record unpredictable events. A tube with a shutter grid following the photocathode has been made and used successfully as a high-speed shutter tube (Mandel 1961, Newth and Walters, private communication).

3.4. *The Channelled Secondary-emission Image Multiplier*

This proposal (McGee 1953), very simple in principle, has been found very difficult to put into practice. Basically the idea is to divide the image into a large number of picture points and to pass the electrons liberated by the light from each picture point through a minute, separate photomultiplier. The multiplied output electron currents from the multipliers are then combined to produce a fluorescent image or a picture signal.

The problem is to devise dynodes of such a geometrical structure that they can be miniaturized sufficiently to give reasonable image definition, can be aligned to

give accurate registration from one dynode to the next, and also give adequate electron multiplication per stage.

It should be said at once that no successful solution to this problem has yet been achieved, but in the opinion of the author it is of such potential importance that it is worth while to give a brief outline of the present state of research.

Active work on this problem has been carried out by Burns and his colleagues of the Chicago Midway Laboratories (Burns and Neumann 1960) and in the author's laboratories at Imperial College (McGee, Flinn and Evans 1960).

The structure adopted by Burns is shown in Fig. 17. This shows three dynodes each consisting of a metal mesh, the bars of which have a trapezoidal cross section. Thus the mesh apertures are cones of square cross section. The

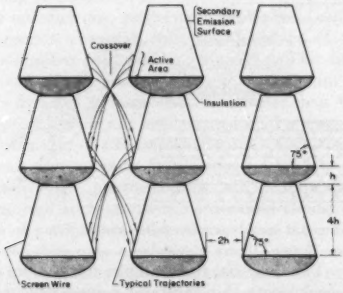


Fig. 17. Trapezoidal screen shape and typical electron trajectories in the channelled image intensifier (*after* Burns and Neumann 1960).

surface of the mesh with the smaller apertures is coated with an insulating layer of, say, MgF_2 or insulating enamel, so that when the meshes are assembled as shown in the diagram they are well insulated from one another. In this device use is made of secondary emission from the surface upon which the primary electrons are incident and hence an electron energy of only a few hundred electron volts is required per stage. This is an order of magnitude lower than the inter-stage potential required in the multi-stage intensifiers described in § 3.3. However, the dynodes must be in close proximity, preferably touching, to prevent electrons straying into neighbouring channels. Hence the field gradient across the inter-dynode insulation will become very large as the dynode dimensions are reduced.

The optimum shape of the apertures in the dynodes was arrived at by tests with a two-dimensional rubber model and then checked by observing the trajectories of actual electrons in a large scale model. The electrostatic field between three successive dynodes was so adjusted that the field at the surfaces of the apertures of the middle dynode is such that electrons liberated from the surface of approximately its lower half find themselves accelerated on to the next dynode. On the upper half of the surface slow secondary electrons would be suppressed but the electrons from the lower half of the previous dynode apertures are directed to the

lower part of the next aperture. Hence the conditions required in any electron-multiplying dynode system are met. Burns claims that about 75% of the secondaries liberated in each dynode are accelerated on to the next. This factor, with the secondary emission coefficient of the surface of the dynodes, determines the gain per stage.

The surface of the apertures of the dynodes must be prepared in such a way as to ensure a high secondary-emission coefficient. It is not easy to satisfy this criterion as well as all those involved in the structure and assembly of such a dynode system. Usually the dynode mesh is made of a metal which is not itself a good secondary emitter, and some layer must be applied to its surface to produce good

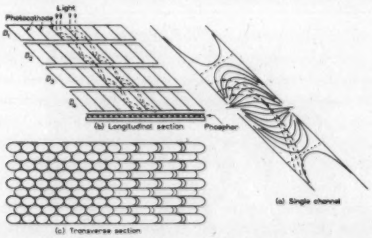


Fig. 18. Channelled image intensifier (*after* McGee, Flinn and Evans 1960).

secondary emission. This may be, for example, magnesium oxide or antimony–caesium formed by evaporating thin layers of magnesium or antimony and subsequently oxidizing or treating with caesium. Both processes have proved to be difficult to do reliably, and this is probably the greatest practical difficulty involved in the making of this device.

Another difficult problem in the design of this device is the introduction of the primary photoelectrons into the first stage so that a large proportion of them produce secondary electrons that pass on to the next dynode.

The most favoured design in the author's laboratory is that illustrated in Fig. 18 (McGee, Flinn and Evans 1960). In this design the dynodes are made by bunching and soldering together cylindrical metal tubes. These are then sawn into short lengths by cutting in planes about 55° to the axis of the tubes. These slices are then reassembled in exactly the same orientation to one another as they had in the original block but are spaced a small distance apart to provide insulation. Now the individual channels can be aligned with great accuracy, which is a crucial problem in the construction of such a system of electrodes.

In Fig. 18, part (*b*) shows a longitudinal section of such an electrode assembly while (*c*) shows a transverse section of the same structure. Part (*a*) shows two adjacent dynode cells and the approximate equipotential lines mapped on a section through the axis. The dotted lines show the approximate paths of slow electrons liberated from the lower half of the upper dynode and it is seen that they land on the surface of the lower half of the lower dynode. It will be fairly evident that correct operation of these dynodes will depend on having the correct relationship

between diameter and length of the tube and the angle at which they are cut. The diagrams illustrate that the conditions for electron multiplication have been provided and measurements on large scale dynodes have shown that the extraction efficiency of secondary electrons is quite good.

This structure lends itself to a relatively simple input dynode D_1. This is made rather shorter than the multiplying dynodes D_2, D_3, etc., and the inner surfaces of the cylinders, as seen from above, are made either photosensitive or good secondary emitters. In the former case the optical image is focused directly onto the end of the dynode while in the latter an electron image is focused in the same plane. In both cases, slow electrons, photo- or secondary-electrons, are liberated in an area of the cylinder from which they can be accelerated onto the lower half of the next multiplying dynode.

It is clear that if all electrons, as they are multiplied, can be prevented from straying from their correct channels, then the definition of the final image will be determined by the number of channels. Since it will be difficult to make these channels of very small dimensions, say less than 0·25 mm diameter, and also not easy to make the dynodes of large area, say greater than 4 in. × 4 in., it seems probable that this type of device will not be capable of giving very high definition, probably not better than that of a television picture. However, if a multiplication factor comparable with that common in electron multipliers can be achieved, of the order of 10^6, then single photoelectrons will be reproduced as visible and photographable events and such an instrument would almost certainly have many applications. It has the important advantages, from the point of view of some applications, that it requires only a relatively small overall working potential, for example less than 10 kv, and this need not be stabilized very precisely as it does not affect the image focus. Also it requires no focus coil and focusing current. These considerations suggest that such a device would be useful for image intensification in a stabilized earth satellite where power consumption and weight must be kept to a minimum and strong magnetic fields are undesirable.

§4. Positive Feedback Image Intensifying Systems

4.1. *General Considerations*

Until quite recently the possibility of operating a single-stage image intensifier with positive optical feedback from the image on the phosphor screen to that on the photocathode was being actively investigated (Roberts 1960, Perl and Jones 1960). This has largely ceased as a result of the success of multi-stage image intensifiers described above. However, it is of sufficient interest to warrant a brief account here.

The method employed is illustrated in Fig. 19. The image tube 1 is operated with uniform electrostatic accelerating and magnetic focusing fields, so producing an upright image of unity magnification on the phosphor screen. The mirror 7 has an aperture 8 through which the lens 2 forms the primary image on the photocathode of the tube. The lens 3 forms an image of the intensified image on the phosphor screen onto a photographic plate through the aperture 9 in the mirror 4. However, a proportion of the light from the phosphor is intercepted

by the mirror 4 and reflected via the mirror 5, a field lens 10 and further mirrors 6 and 7 to be focused as far as possible in accurate registration with the primary image.

If now the quantum gain of the tube is A and a fraction F of the output light is fed back to the photocathode image, then we have a loop gain of $AF = G$. If $G > 1$, then an image flashed on to the photocathode will produce an image on the phosphor screen of increasing intensity until it is limited by the overloading of the tube, deliberate interruption of the power supplies or feedback path of the tube. If $G < 1$, then the image will fade with time.

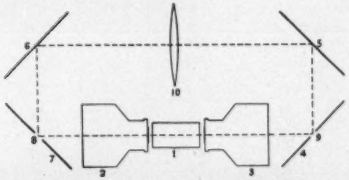


Fig. 19. Single-tube regenerative image intensifier (*after* Perl and Jones 1960).

If a signal of I photons is flashed on the photocathode and we assume no time lag in emission of light by the phosphor screen, after the signal has passed through the tube C times the light output L from the screen will be given by $L = IAG^{C-1}$.

In practice the light is emitted by the phosphor according to a decay characteristic which may be approximately exponential, but is usually a complex sum of several exponential decays of very different time constants. The most efficient phosphor for such an intensifier tube ZnS:Ag, or P-11, has a half-value period of the order of 0·1 msec and hence light from the first fluorescent image will continue to be fed back to the photocathode for about 1 msec until it has completely faded. On the other hand, a considerable proportion of the total light emission will take place in the first few microseconds after the arrival of the exciting electrons. This is long compared with the feedback time due to the time of travel of the light, approximately 10^{-9} sec, and the electrons, approximately 10^{-8} sec. The result of this is that a whole series of successive images grow up on the phosphor screen before the original image has faded. This complication could be avoided only by using a phosphor of decay time shorter than the optical plus electron feedback time. While this might be achieved by stretching both the optical and the electron path to a maximum and choosing the phosphor with the shortest decay period, it is found that such phosphors are very inefficient and the result would be a serious decrease in loop gain G.

4.2. *Single-tube System with Accurate Registration*

The obvious way to attempt regeneration is to achieve accurate registration between the primary optical image on the photocathode and the secondary image reflected back to it from the phosphor screen. If this can be near perfect, then with only a small loop gain, but a sufficiently large number of cycles, a large image intensification, without loss of definition, could be achieved. The regenerative action must be interrupted periodically when the intensification has reached an

adequate amount, either by interrupting the light feedback or gating the electron stream in the tube for a period comparable with the time of the decay of the phosphor. However, until recently, it has not been practicable to achieve either large gain or adequate fidelity of image formation to render this possible. It is obvious that if there is serious loss of definition or loss of geometrical fidelity, as between the optical and the electron image, then the definition of the regenerated image will become progressively worse the larger the number of cycles necessary, that is, the smaller the value of F. In the past, when the image tubes available were mainly electrostatic focusing, this problem of registration appeared insuperable because of the poor image resolution at points a little off the axis and the serious pin-cushion image distortion. It is probable that, with the improvement in both gain, image definition and geometrical fidelity, reported above (§ 2.3), fine registration in a regenerative system could be achieved. However, in the meantime, the success of the multi-stage intensifier tubes (§§ 3.2 and 3.3) has largely eliminated the necessity for the regenerative method.

4.3. *Systems without Registration*

In an attempt to avoid the difficulties of accurate registration the possible use of complete non-registration of primary and fed-back images was considered. If the loop gain G is reasonably large, say even 10 times, then the $(n+1)$th image will emit 10 times the light emitted by the nth image. Hence by suitable photography the $(n+1)$th image might be recorded with little interference from the nth. However, this overlooks the effect of the decay time of the phosphor being long compared with the feedback time. It can be shown (Roberts 1960) that the brightnesses at any time of the series of non-registered images on the phosphor screen are related by very similar equations to those relating to the number of atoms present in a series of genetically related radioactive elements having the same half-life and a multiplication factor G introduced at each transformation. Thus it works out that at any instant the nth image is not very different in brightness from the $(n-1)$th or $(n+1)$th image. Thus it is not practicable to photograph one particular image and suppress the others.

A method to avoid this difficulty was proposed by Perl and Jones (1960) which required the use of two image tubes through which the image could be passed in successive intervals. When the first tube is operative the second is inoperative and vice versa. Thus the signal moves round the loop in a discontinuous fashion and an infinite series of superimposed images is avoided. There is a last and brightest image. This method also appears to have been overtaken by the successful development of the multi-stage tubes.

§ 5. Image Intensifier Tube Applications

Brief outlines will now be given of the uses, or possible uses, of image tubes in the fields of astronomy, nuclear physics, spectroscopy and radiology.

5.1. *Astronomy*

In modern observational astronomy there appear to be two main problems (McGee 1955, 1956). These are: first, to increase the efficiency of recording

of light photons collected by optical instruments above that achieved by direct photography and second, to detect faint objects against the bright, night-sky background.

The first of these problems may be considered to relate to small and medium telescopes for which the practicable exposure time of the best photographic plate is not limited by the fogging of the emulsion due to sky background. For this type of instrument what is required is an ideal image intensifier which would record about 100 times larger proportion of the incoming photons without losing image definition, geometrical accuracy, gradation or introducing spurious background which would limit the exposure or impair image gradation. Given this quality of performance, an image intensifier would improve the performance of a 20 in. reflector telescope to compare with that of the 200 in. using normal photography only.

To reach this performance an image intensifier should resolve approximately 40 line-pairs per mm over a field of approximately 40 mm diameter. Geometrical accuracy must be very good and it must be possible to make an exposure with it for $\frac{1}{2}$ to 1 hour without appreciable fogging of the emulsion due to spurious tube background.

Of the image tubes described above none quite meets this specification. Most fail in one or two aspects, but the ideal device does not seem to be beyond the possibilities of further development. It should be pointed out that the figure of merit of any such device is approximately proportional to the product of its light gain, the square of its resolution in line-pairs per mm and the effective area of the image, assuming background to be negligible.

The use of such an image intensifier for photography of the heavens by the 200 in. telescope would not give a worth-while advantage since, although it would enable the sky-background limit to be reached more quickly, perhaps in $\frac{1}{2}$ minute instead of 30 minutes exposure, it would not increase the discrimination of the telescope of faint objects against the sky background, which already limits its range. If, however, in addition to the attributes listed above, the image intensifier could enable us to record more information from an image, for example, by recording on a fine-grained emulsion by the methods described in §§ 2.4, 2.5 and 2.6, or on a large area of light-sensitive emulsion by means of the devices described in § 3, then it should be possible to use a device to improve the detection of objects of very small contrast for the following reason.

The possibility of detecting an image of an object in a pattern of image and background formed by random discrete dots is given by expression (2). This shows that the larger the number of discrete events that can be recorded the smaller the contrast of small objects that can be detected. In fact, if we could record 100 times as many photons per unit area of an image, as appears possible on fine-grained nuclear emulsions, then the contrast step that we can detect should be $\sqrt{100}$ turns or about 10 times smaller. This represents 2·5 stellar magnitudes. However, assuming a gain by a factor of 100 times in the efficiency of recording photons, it would require the same exposure time as is now used with direct photography, but the limit of detectability would be 2·5 magnitudes fainter.

Probably the major part of astronomical information is derived from spectra and again, in many cases, the amount of light available is the limiting factor. In

this case sky background is not generally an important limiting factor. Moreover, in this case a long, narrow rectangular field of observation and recording is satisfactory. Hence in the case of those devices described in §§ 2.2, 2.3 and 2.4 with thin mica windows this should simplify the construction of such tubes and allow thinner mica to be used with a consequent improvement in sensitivity and resolution.

5.2. *Nuclear Physics*

Two observations for which image intensifier tubes are essential are: first, the recording of fast nuclear particle tracks in transparent blocks of phosphor, thus giving information about nuclear interactions similar to that obtained from the Wilson cloud chamber or the bubble chamber due to Glaser and second, the detection of the Čerenkov radiation from fast particles in a transparent, dense medium from which their velocities can be determined and their masses deduced.

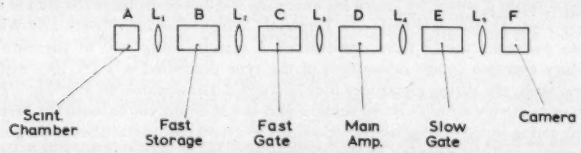


Fig. 20. Scintillation chamber, image tubes and camera for recording fluorescent tracks of nuclear particles.

In both these cases the amount of light produced by the particle is very small, far less than the minimum detectable by a photographic plate. In fact it is only those image intensifiers described in §§ 3.2 and 3.3 with very large light gain, such that single photoelectrons can be recorded as distinct dots, that are suitable for this purpose. Another important requirement of the observing device used for this purpose is that it must be capable of being gated ' on ' and ' off ' at very high speed, say in a few nanoseconds (mμsec), without serious loss of electrons in the ' on ' state. In fact it is this feature, combined with its extreme sensitivity to light, that gives this device its unique role in nuclear physics observations since by means of it we are able to select and record events of significance after they have happened from vast numbers of uninteresting events.

The general arrangement for the ' scintillation chamber ' for particle track detection is shown in Fig. 20. First, the scintillation chamber A is a block of transparent phosphor such as NaI or CsI if high efficiency but rather slow response, of the order of 10^{-6} sec, is required, or anthracene or plastic phosphor if very fast response, of the order of 10^{-8} sec, is necessary and rather lower efficiency can be tolerated.

The light from the track of a nuclear particle passing through the block of phosphor is focused by an optical system L_1 onto the photocathode of the first stage of an image intensifier and the photoelectrons liberated are focused onto a phosphor screen, at the other end of this section, which has a fast decay period, say 10^{-6} to 10^{-7} sec, and hence an image of a particle track will be held on this screen for a period of time of this order and will then fade out. Whether or not this image will be passed on to the main amplifier section D will be determined

by the fast gate C which can be opened or closed at high speed, and the decision to open this gate can be made on the basis of information provided by a signal from the scintillation chamber, for example from two photomultipliers in coincidence. Thus the fast gate can be kept closed until some event of significance happens, thus keeping the phosphor screens of the main amplifier, which must be more efficient but slower in decay, free from an excess of unwanted tracks and spurious background. Once the image has passed the fast gate it is intensified in the main amplifier and passes the slow gate E, which may be the shutter of the camera lens L_5 by which the final image is photographed on the film F.

The sections B, C and D need not be separate but may be built into two or even one tube. In the first reported successful system of this type, due to Zavoiskii *et al.* (1955, 1956), all the functions were carried out in a single tube. Subsequently, systems which have been able to detect tracks of particles have been built up out of several tubes coupled by lenses by Reynolds and his co-workers (Reynolds 1959, Condon 1960, Perl *et al.* 1960). In Imperial College, London, Newth and Walters (private communication) have constructed a system using two of the thin-film secondary emitting image intensifiers of the type described in § 3.3, one without, and one with, the gating electrodes outlined in § 2.9 and shown in Fig. 15. These both had relatively slow ZnS: Ag screens and hence could not be used for very fast discrimination but are being used successfully to detect cosmic particles ; the passage of a particle and the opening of the electron gate being signalled by a coincidence signal from two photomultipliers, one above and one below the phosphor block.

The quality of the particle tracks published to date leaves much to be desired if useful nuclear physics experiments are to be done with this method. The features most in need of improvement are the diameter and efficiency of the primary photocathode and the definition of the final image. The required improvements seem to be reasonably within reach.

Čerenkov radiation has not yet been detected, but it appears that it should be just about possible to do so even now. An increase by a factor of 3 in photocathode sensitivity and effective diameter, both well within present possibilities, would make detection relatively reliable.

5.3. *Spectroscopy*

As in astronomical spectroscopy, so in other applications of spectroscopy, there are situations in which the amount of light available is a serious limitation and in these problems any one of the image intensifiers described above may be used. However, in the vacuum ultra-violet, not only may the light intensity be small but there is the additional problem of passing it through glass or quartz windows into a detecting device. Even if the image intensifier is placed inside the chamber in which the ultra-violet is generated this still remains a problem since in general it is necessary to let air into these chambers at times and this would kill all image intensifiers using conventional photocathodes unless enclosed in a vacuum-tight envelope.

For vacuum ultra-violet, say below 2000 Å, a clean gold or other metal surface is an efficient photocathode and this surface is not impaired by exposure to clean, dry air. Hence such a surface could be used as the photocathode and the resulting photoelectrons could be detected by one or other of the methods described in §§ 2.2, 2.3, 2.4 and 3.3. In the first three of these methods the electrons would be

accelerated and focused on a thin mica window in the wall of the chamber and the recording be made by bringing a film into contact with the outer surface of the window.

In the fourth method the electrons would be multiplied at a series of thin-film dynodes and the intensified image focused onto a phosphor screen to produce an image which could easily be photographed from without the main chamber. The cascade phosphor–photocathode intensifier screen could not be used for this purpose but the thin composite films would be satisfactory since they remain efficient for long periods when exposed to air provided it is clean and dry.

5.4. *Radiology*

It is of great importance to be able to detect x-ray patterns with such sensitivity that the image information is limited only by the shot noise of the x-ray quanta. This is important in scientific observation simply to enable observations to be made more quickly and accurately than before. It is of great importance in medical radiology in order to allow x-ray examinations to be made with the minimum radiation dosage to the patient.

The availability now of an image intensifier which can amplify the light of an incident image by approximately 10^5 times and has a photocathode which can have about 10% quantum efficiency makes it possible to couple a conventional x-ray-sensitive phosphor screen by means of a lens with the cathode of such an intensifier tube so as to release at least one photoelectron for each x-ray photon absorbed in the phosphor screen. The photocathode of the intensifier tube available at present, of the type described in § 3.3, is only 1 in. in diameter, and hence to have a reasonable size of phosphor screen, say 4 in. diameter, the fluorescent image must be reduced by a factor of four to the photocathode. Thus only 0·5 to 1·0% of the light produced by an x-ray quantum absorbed in the phosphor screen will reach the photocathode. However, we may expect between 3000 and 5000 photons of blue-green light to be liberated by each absorbed quantum (Morgan and Sturm 1949, Hay 1960). Thus we may expect an average of between 0·75 and 5 electrons to be liberated from the photocathode for each x-ray quantum absorbed in the phosphor screen. It is important that this number be kept as large as possible in order to minimize the statistical fluctuations that will be introduced at this point, which become larger the smaller this transfer factor. If the photocathode efficiency can be improved and the area of the cathode increased, it seems reasonable to hope that this figure will soon be not less than 5. To achieve an output signal/noise ratio of say 10:1, at least 100 x-ray quanta, per picture point per frame, must contribute to the picture signal. Additional noise will arise from the fluctuations in the above transfer factor and this will become less the larger the factor ; above about 5 it will be negligible.

It is the usual practice of radiologists to examine a patient first with a large area screen, say 20 cm × 30 cm, and then to reduce the area, perhaps to 8 cm × 10 cm, for closer examination. It is fortunate that in the former case, where the collection of light by the objective lens will be less than in the latter, there will be a much larger number of effective x-ray quanta from which to collect light, while in the latter case there will be fewer quanta to produce light but about five times as many photons can be collected from each absorbed x-ray quantum. Hence, the noise

due to the primary x-ray quanta will increase, assuming no increase of x-ray flux, but that additional noise due to the conversion of x-ray quanta to photoelectrons will decrease.

Since the image can be intensified to a very high level in the intensifier tube it can be viewed directly with a magnifying eye-piece or it can be coupled to a television camera and the image can be reproduced as a television display. Because of the high brightness level of the intensifier output image, it is possible to transfer sufficient light to the television tube to obtain a picture signal in which the noise addition from the camera is negligible.

Acknowledgments

I should like to thank the various authors and their publishers for permission to reproduce their diagrams. I also wish to thank my colleagues of the Instrument Technology Section of the Physics Department of Imperial College for their contributions, both in experiment and discussion, to this work.

References

Bril, A., and Klasens, H. A., 1952, *Philips Res. Rep.*, **7**, 401.

Brüche, E., and Busch, H., 1937, *Beiträge zur Elektronenoptik* (Leipzig : Barth), p. 107.

Burns, J., and Neumann, M. J., 1960, *Advances in Electronics and Electron Physics*, Vol. XII, *Proceedings of Symposium on Photoelectronic Image Devices, Imperial College, London*, 1958 (London, New York : Academic Press), p. 97.

Butslow, M. M., Zavoiskii, E. K., Kalinyak, A. A., Nikonov, V. V., Prokofiev, V. V., and Smolkin, G. E., 1958, *Dokl. Akad. Nauk*, **121**, 815.

Condon, P. E., 1960, *Advances in Electronics and Electron Physics*, Vol. XII, *Proceedings of Symposium on Photoelectronic Image Devices, Imperial College, London*, 1958 (London, New York : Academic Press), p. 128.

Davis, E. J., 1959, *M.Sc. Thesis*, University of London.

Eckart, F., 1937, *Elektronenoptische Bildwandler und Röntgenbildverstärker* (Leipzig : Barth), p. 78.

Engstrom, R. W., 1960, *R.C.A. Rev.*, **21**, 184.

Ennos, A. E., 1953, *Brit. J. Appl. Phys.*, **4**, 101.

Ford, W. K., Fredrick, L. W., Hall, J. S., Houck, T. E., and Baum, W. A., 1960, *Astr. J.*, **66**, in the press.

Hall, J. S., Ford, W. K., Jr., and Baum, W. A., 1960, *Advances in Electronics and Electron Physics*, Vol. XII, *Proceedings of Symposium on Photoelectronic Image Devices, Imperial College, London*, 1958 (London, New York : Academic Press), p. 21.

Hay, G. A., 1960, *Advances in Electronics and Electron Physics*, Vol. XII, *Proceedings of Symposium on Photoelectronic Image Devices, Imperial College, London*, 1958 (London, New York : Academic Press), p. 363.

Hiltner, W. A., 1956, *Symposium on Present and Future of Telescope of Moderate Size* (Philadelphia : University Press), p. 11.

Holland, L., 1959, *Nature, Lond.*, **181**, 1451.

Kollath, R., 1941, *Ann. Phys., Lpz.*, **39**, 59 ; 1947, *Ann. Phys., Lpz.*, **1**, 357.

Krassovsky, V. I., and Shcheglov, P. V., 1958, *Trans. I.A.U.*, **10**, 145.

Lallemand, A., 1936, *C.R. Acad. Sci., Paris*, **203**, 243, 290.

Lallemand, A., and Duchesne, M., 1951, *C.R. Acad. Sci., Paris*, **233**, 305 ; 1952, *C.R. Acad. Sci., Paris*, **235**, 503 ; 1954, *C.R. Acad. Sci., Paris*, **238**, 335.

Lallemand, A., Duchesne, M., and Walker, M. H., 1960, *Publ. Astron. Soc. Pacific*, **72**, 76.

Lallemand, A., Duchesne, M., Wlérick, G., Augarde, R., and Dupré, M. F., 1960, *Ann. Astrophys.*, **23**, 320.

LUBSZYNSKI, H. G., 1935, British Patent No. 457493, 30/5/35.
LUBSZYNSKI, H. G., and RODDA, S., 1934, British Patent No. 442666, 12/5/34.
MCGEE, J. D., 1937, British Patent No. 504927, 28/10/37 ; 1953, British Patent No. 790416, 5/6/53 ; 1955, *Symposium on Astronomical Optics, Manchester, April* (edited by Z. Kopal) (Amsterdam : North Holland), p. 21 ; 1956, British Patent No. 9796/56, 29/3/56 ; 1956, *Present and Future of Telescope of Moderate Size* (edited by F. Bradshaw Wood) (Philadelphia : University Press), p. 35 ; 1959, *Image Intensifier Symposium, U.S.A.E.R.D.L., Fort Belvoir* 1958 (Washington, D.C. : U.S. Department of Commerce, O.T.S., 151813, 500).
MCGEE, J. D., FLINN, E. A., and EVANS, H. D., 1960, *Advances in Electronics and Electron Physics*, Vol. XII, *Proceedings of Symposium on Photoelectronic Image Devices, Imperial College, London*, 1958 (London, New York : Academic Press), p. 87.
MCGEE, J. D., and LUBSZYNSKI, H. G., 1939, *J. Instn Elect. Engrs*, **84**, 468 ; 1940, British Patent No. 542245, 27/6/40.
MCGEE, J. D., and WHEELER, B. E., 1961, *Royal Photographic Soc. Symposium, London* 1960, *J. Photographic Sci.*, **9**, 106.
MANDEL, L., 1955, *J. Sci. Instrum.*, **32**, 405 ; 1959, *Optica Acta*, **6**, 354 ; 1961, *Proceedings of 5th International Congress on High Speed Photography, Washington*, 1960 (in the press).
MILLER, R. H., HILTNER, W. A., and BURNS, J., 1956, *Astrophys. J.*, **123**, 368.
MORGAN, R. H., and STURM, R. E., 1949, *Amer. J. Roentgenol.*, **62**, 617.
MORTON, G. A., and FLORY, L. E., 1946, *Electronics*, **19**, 112.
ORVIN, L. J., 1934, British Patent No. 445156, 31/5/34.
PERL, M. L., and JONES, L. W., 1960, *Advances in Electronics and Electron Physics*, Vol. XII, *Proceedings of Symposium on Photoelectronic Image Devices, Imperial College, London*, 1958 (London, New York : Academic Press), p. 153.
PERL, M. L., JONES, L. W., NEWSOME, R., and LAI, K., 1960, *Report on Office of Naval Research Contract* No. Nonr–1224 (23) (University of Michigan).
PHILIPS GLOEILAMPENFABRIEKEN, 1928, British Patent No. 326200, 5/11/28.
REYNOLDS, G. T., 1959, *Proc. C.E.R.N. Conference, Geneva*, p. 576.
ROBERTS, A., 1960, *Advances in Electronics and Electron Physics*, Vol. XII, *Proceedings of Symposium on Photoelectronic Image Devices, Imperial College, London*, 1958 (London, New York : Academic Press), p. 135.
ROSE, A., 1948, *Advances in Electronics* (edited by L. Marton), Vol. I, p. 131 (New York : Academic Press).
SHCHEGLOV, P. V., 1960, *Astron. Zh.*, **37**, 586.
SOMMER, A. H., 1955, *Rev. Sci. Instrum.*, **26**, 725.
STERNGLASS, E. J., 1956, *Rev. Sci. Instrum.*, **26**, 1202.
STOUDENHEIMER, R. G., 1960, *Advances in Electronics and Electron Physics*, Vol. XII, *Proceedings of Symposium on Photoelectronic Image Devices, Imperial College, London*, 1958 (London, New York : Academic Press), p. 41.
TEEVES, M. C., and TOL, T., 1952, *Philips Tech. Review*, **14**, 37.
VOLKOV, I. V., YESIPOV, V. F., and SHCHEGLOV, P. V., 1959, *Dokl. Akad. Nauk*, **129**, 288.
WACHTEL, M. M., DOUGHTY, D. D., and ANDERSON, A. E., 1960, *Advances in Electronics and Electron Physics*, Vol. XII, *Proceedings of Symposium on Photoelectronic Image Devices, Imperial College, London*, 1958 (London, New York : Academic Press), p. 59.
WILCOCK, W. L., EMBERSON, D. L., and WEEKLEY, B., 1960, *Nature, Lond.*, **185**, 370 ; 1960, *Trans. Instn Radio Engrs*, NS–7, 126.
YESIPOV, V. F., 1960, *Astron. Zh.*, **37**, 588.
ZACHAROV, B., and DOWDEN, S., 1960, *Advances in Electronics and Electron Physics*, Vol. XII, *Proceedings of Symposium on Photoelectronic Image Devices, Imperial College, London*, 1958 (London, New York : Academic Press), p. 31.
ZAVOISKII, E. K., BUTSLOV, M. M., and SMOLKIN, G. E., 1955, *Dokl. Akad. Nauk*, **100**, 241 ; 1956, *Dokl. Akad. Nauk*, **111**, 996.

FERRIMAGNETISM

By W. P. WOLF
Clarendon Laboratory, Oxford

CONTENTS

	PAGE
§ 1. Introduction	213
§ 2. Ferrimagnetic compounds	217
2.1. General considerations	217
2.2. Crystal structures	220
2.3. Methods of preparation	226
2.3.1. Polycrystals	227
2.3.2. Single crystals	227
§ 3. Magnetostatic properties	229
3.1. High temperature susceptibility	229
3.2. Spontaneous magnetization	232
3.2.1. The magnetic moment at $T=0°K$	232
3.2.2. Temperature dependence of magnetization	236
3.3. Susceptibilities below the Néel point	240
3.4. Crystalline anisotropy	243
3.4.1. Experimental methods	245
3.4.2. Theory	245
3.4.3. Macroscopic effects	250
3.5. Magnetic annealing	251
3.6. Magnetostriction	252
§ 4. Thermal properties	253
4.1. Specific heat	253
4.2. Thermal conductivity	255
4.3. Magnetocaloric effects	256
§ 5. Microwave properties	257
5.1. Dynamic susceptibility	257
5.2. Ferrimagnetic resonance: Introduction	260
5.3. Effective g-factor	261
5.4. Effect of anisotropy on resonance	266
5.5. Resonance near compensation points	270
5.6. Non-uniform resonance modes	274
5.7. Line widths	275
5.8. High power effects	281
5.9. Dynamic effects in unsaturated samples	285
5.9.1. Multiple microwave resonances	285
5.9.2. Domain wall motion	285
5.9.3. Initial permeability	286
§ 6. Experiments with neutrons	287
6.1. Neutron diffraction	287
6.2. Inelastic neutron scattering	289
§ 7. Optical properties	290
7.1. Absorption and Faraday rotation	290
7.2. Observation of domains	292
§ 8. Recent developments	292
8.1. Complex spin configurations	292
8.2. Sub-lattice magnetization	293
8.3. Spin waves and non-linear effects	294
Acknowledgments	294
References	295

Abstract. This is a review of the physical properties of ferrimagnetic materials. These are substances which exhibit spontaneous magnetization due to non-parallel alignment of their atomic magnetic moments, and which at the same time are generally poor conductors of electricity. Some of their characteristic properties arise from the multidirectional alignment, while others arise from the fact that it is possible to carry out experiments at very high frequencies without interference from eddy currents. Due to the considerable difficulties in preparing suitable samples, detailed understanding of many properties has only recently become possible, and a large number of ' spurious ' effects have been detected in earlier measurements. This review will therefore concentrate on the kinds of experiments and theories which have been developed, rather than a summary of individual results.

§ 1. Introduction

The word ferrimagnetism was coined by Néel (1948 a) to describe the properties of those substances which below a certain temperature exhibit spontaneous magnetization arising from a non-parallel alignment of atomic magnetic moments. In his original paper Néel envisaged a partitioning of the moments into two sub-lattices, which because of their mutual interaction are aligned anti-parallel to each other, thus producing a total magnetic moment equal to the difference between their individual magnitudes. This difference can arise in several ways, and some of these are shown diagrammatically in Fig. 1. Fig. 1(*a*) shows the case in which there are different numbers of similar magnetic moments in the sub-lattices. In some respects this arrangement superficially resembles that of a normal antiferromagnetic substance with unequal partitioning of the sub-lattices, but many of the properties of such a substance are similar to those of a ferromagnet with a reduced number of magnetic moments. Fig. 1(*b*) shows the case of equal numbers of dissimilar moments. The dissimilarity can arise either because the magnetic ions are chemically different, or because a different local environment leads to different effective magnetic moments in ions having the same spin. So far only the first of these possibilities has been observed. A third type of arrangement is shown in Fig. 1(*c*). This is the case originally considered by Néel and it represents a very large number of substances, including most of the ferrites. In this arrangement one sub-lattice contains two different types of magnetic moment, one of which also occurs in equal numbers on the second sub-lattice, so that the net effect is that of just the one type of moment. An extension of this case arises when the second sub-lattice also contains two types of moments, as shown in Fig. 1(*d*).

With this simple idea Néel was able to explain quantitatively the observed magnetic moments of a large number of substances, and we shall consider some specific examples in § 3.2. An extension of the Néel theory has been given by Yafet and Kittel (1952), who pointed out that if there are strong interactions between the ions *within* a given sub-lattice, in addition to the interaction between the sub-lattices, then a triangular arrangement such as shown in Fig. 1(*e*) will have a lower energy than the uniaxial Néel type arrangements. More recently Kaplan (1959) has shown that even triangular arrangements are not stable in certain cases, for which more complicated ' screw-type ' arrangements may then occur (see also Yoshimori 1959, Villain 1959, Bertaut 1960). Obviously such situations can be extended to a large number of variants, involving various three-dimensional multi-spin-axis arrangements. So far the experimental evidence for

such arrangements is somewhat scanty and controversial, and we shall consider this in more detail in § 3.2.

An alternative type of multi-spin-axis arrangement, which does not seem to have been considered before, arises when there are strong local anisotropy forces tending to hold each spin in a particular direction relative to the crystal sub-lattice. Consider, for example, the simple case of a cubic crystal in which one sub-lattice consists of three types of moment, each of which is bound by some anisotropy forces to one of the three cube axes. Interaction with a second sub-lattice, which we may for simplicity assume to be free, then produces a compromise arrangement such as shown in Fig. 1(*f*). So far there is no direct experimental evidence for such an arrangement, but it seems very likely that some ferrimagnetic rare earth compounds have this type of arrangement.

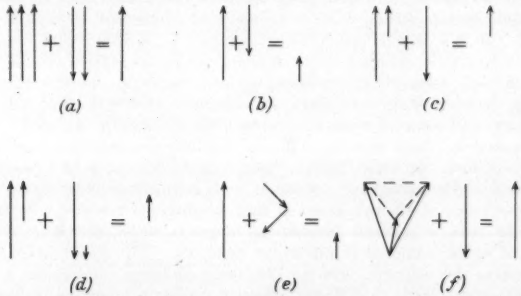


Fig. 1. Illustration of six simple sub-lattice arrangements which can give rise to a spontaneous ferrimagnetic moment.

In general there is nothing to limit the number of sub-lattices and their direction of alignment, and it is important to realize that the simple arrangements we have discussed are only a first approximation in most real crystals. However, for the explanation of many phenomena the simple Néel theory is often quite adequate.

In addition to explaining the magnetic moments in the completely ordered state, Néel was able to calculate from his model the variation of magnetic moment with temperature in terms of the various interactions between the atomic moments. These interactions he grouped according to the sub-lattices into an average (A–B) coupling, tending to make A and B antiparallel, together with average intra-sub-lattice forces (A–A) and (B–B). Using a molecular field theory he was then able to calculate the variation of magnetization for various ratios of the interactions, and predicted the curves shown in Fig. 2. All these have since been observed experimentally. Néel's theory also predicted several other types of (M_s, T) curves which cannot occur in practice since they require $\partial M_s/\partial T$ to be finite at $T = 0°\text{K}$, in contradiction to the third law of thermodynamics. The difficulty was removed by Yafet and Kittel who pointed out that these curves correspond to large intra-sub-lattice couplings, in which case a triangular or multi-spin-axis arrangement will have a lower energy.

The anomalous shapes of the (M_s, T) curves arise from the fact that M_s is the resultant of two or more components, each of which approaches its own saturation value as a different function of temperature (see for example Fig. 2(*e*)). This gives rise to the possibility of a ' compensation point ', a temperature T_c at which the magnetic moments of two sub-lattices are just equal and opposite, as shown in Fig. 2(*e*). This is a typically ferrimagnetic phenomenon and it gives rise to some striking microwave resonance effects which are discussed further in § 5.5. However, even when there is no compensation point the presence of more than one inequivalent magnetic sub-lattice can give rise to collective properties, such as anisotropy or microwave resonance, very different from those of a single ferromagnetic system.

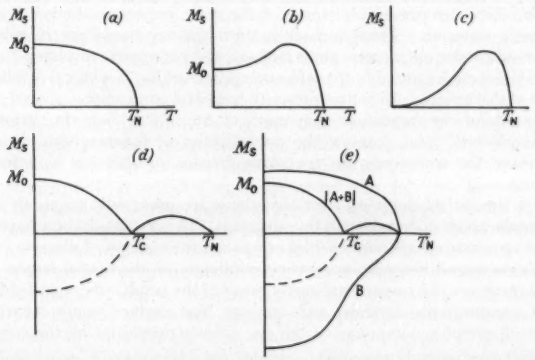


Fig. 2. (*a*), (*b*), (*c*), and (*d*). Four types of spontaneous magnetization–temperature curves predicted by Néel's (1948) theory. (*e*) Illustrates the origin of a compensation point T_c arising from the cancellation of two sub-lattice magnetizations.

Another very important class of properties, which is quite unrelated to the peculiar sub-lattice structure of ferrimagnetics, arises almost accidentally from the nature of the substances which happen to be ferrimagnetic. It is an empirical fact that almost all substances which are ferrimagnetic are non-conductors of electricity, in contrast to ferromagnetic substances which are almost all metals or very good conductors.† This difference may be due to the distinct nature of the interaction forces which give rise to the two kinds of ordering, one being a force which favours antiparallel alignment, the other parallel alignment. We shall discuss these interactions briefly in § 2.1, but in general we need only note the difference as an

† Strictly speaking they are all really semiconductors (see for example Van Uitert 1956, Jonker 1959), but conductivities are generally so low that for practical purposes they can be considered to be insulators.

Note added in proof : Exceptions to this rule have recently been found for a few unusual substances : Mn_3Ge_x (Yasucochi, Kanematsu and Ohoyama 1960), $Mn_{1-x}Cr_xSb$ with $0{\cdot}02 \leqslant x \leqslant 0{\cdot}2$ (Swoboda, Cloud, Bither, Sadler and Jarrett 1960), Mn_4N and other nitrides (see Takei, Shirane and Frazer 1960 for several references). These are all intermetallic compounds which show non-parallel ordering of their magnetic moments, but for many practical purposes it is still preferable to restrict the term ' ferrimagnetic ' to substances which are effectively insulators.

empirical fact. From a practical point of view there are immense advantages in studying insulators, since with these one can carry out experiments over the entire range of the electromagnetic spectrum, without interference from eddy currents. This enables one to use bulk samples even at high frequencies, whereas in the case of metals one is always restricted by the skin depth. Thus one can study various shape dependent effects, observe the propagation of electromagnetic waves through the magnetic medium, and in favourable cases one can even observe the transmission of optical radiation (see § 7).

In addition to the information which can be deduced from such experiments, high frequency magnetic properties are of very considerable importance in technical applications. In this review we shall not be discussing this very large field, and we mention here, by name only, some of the most important devices which use ferrimagnetic materials. These include radio-frequency transformers, microwave isolators, modulators, circulators, phase shifters, and parametric amplifiers. Details of some of these can be found in the references listed under 'Technical Applications' at the end of this article. In connection with technical applications, special mention must be made of the important early work of Snoek (1947 a) who, prior to the theoretical ideas of Néel, foresaw the potentialities of ferrimagnetic compounds and pioneered the techniques for preparing ferrites suitable for high frequency applications.

From a theoretical point of view also there are great advantages in working with materials which do not contain free electrons. This avoids all the complications from band structure effects, and enables one to use the concept of discrete magnetic ions which are bound ionically, or at worst covalently, in the crystal lattice. Under these circumstances one can apply directly many of the results of the well-developed theory of paramagnetic crystals, and one can also deduce many ferrimagnetic properties of complex compounds from the known properties of the constituent ions.

Unfortunately it is not always easy to decide exactly what the constituent ions are, as there are considerable difficulties in preparing pure samples of accurately known composition. There are several reasons for this and these will be discussed in § 2.3, but we may already emphasize here that the need for really pure and stoichiometric samples is still probably the most urgent problem in ferrimagnetism. In recent years many attempts have been made to prepare single crystals of ferrimagnetic substances and these have met with some success. It turns out however that even single crystals have not always been pure enough to be reproducible in all respects, and for this reason we shall present in this review only a very limited number of experimental data and concentrate more on discussing the *types* of experiments and theories which have been developed for the study of ferrimagnetism. It seems reasonable to hope that the problems of preparing adequate samples will be overcome in the near future.

A possible solution of this problem may come through the development of new types of compounds, rather than through the refinement of existing types. Already much has been achieved along these lines, and the discovery of the garnets (Bertaut and Forrat 1956 a, Geller and Gilleo 1957 a) has been a great step forward. It is significant to note that previous reviews of ferrimagnetism (Snoek 1947 a, Fairweather, Welch and Roberts 1952, Gorter 1954, Smit and Wijn 1954) discuss

almost exclusively ' ferrites ', spinel type compounds containing Fe^{3+}, while today there are well over a dozen different kinds of compounds known. Even now the word ferrite is often used synonymously with ' ferrimagnetic compound ', but this will certainly not be so in the future.

§ 2. Ferrimagnetic Compounds

2.1. *General Considerations*

At sufficiently low temperatures all paramagnetic substances become ordered, but at present it does not seem possible to predict which will become ferrimagnetic. One necessary requirement is an interaction between the magnetic ions which tends to align them into a non-parallel array, but depending on the particular structure this may give rise either to antiferromagnetism or to ferrimagnetism. Almost all known ferrimagnetic compounds have rather complex structures, with at least two crystallographically inequivalent sites for the magnetic ions to provide a basis for the sub-lattices, but there appears to be no fundamental reason for this, since ordering of the type shown in Fig. 1(*b*) could exist even in a simple structure. Conversely some complex compounds are antiferromagnetic, and in general the type of ordering can only be determined by experiment.

The temperature at which long-range order sets in, now generally known as the Néel point T_N, is determined largely by the strength of the interaction between nearest neighbours, though under special conditions of symmetry other factors may also be important (Anderson 1956). For both antiferromagnetic and ferrimagnetic insulators the interaction arises from an indirect exchange mechanism, in which the magnetic ions are coupled through electron transfer with intermediate anions. The concept of indirect exchange was put forward by Kramers (1934) and it has been developed in the form known as *superexchange* by Anderson (1950, 1952, 1959), Van Vleck (1951) and others. Different, but superficially similar indirect exchange mechanisms have been proposed by Slater (1953), Pratt (1955), Goodenough and Loeb (1955), Wickham and Goodenough (1959), Kanamori (1959) and Casselman and Keffer (1960). We shall not discuss these theories here as an excellent review of the general position has been given by Nagamiya, Yosida and Kubo (1955) with reference to antiferromagnetism, while the application to ferrimagnetism has been discussed by Gorter (1954) and by Smit and Wijn (1954, 1959). Details of the calculations are generally very complicated but one can briefly summarize the results of all the theories by a simple interaction Hamiltonian of the form

$$\mathscr{H}_{ij} = J_{ij}\,\mathbf{S}_i.\mathbf{S}_j \qquad \ldots\ldots(2.1)$$

coupling the spins $\mathbf{S}_i$ and $\mathbf{S}_j$ of neighbouring magnetic ions i and j. J_{ij} is a constant which depends strongly on the two metal–anion distances and also on the angle subtended by the two magnetic ions at the anion. The form of this coupling is the same as that of the Heisenberg direct exchange interaction, but the sign of J is generally positive so that the state of lowest energy corresponds to antiparallel alignment. Although the theory explains qualitatively the magnitudes of the interactions required to fit the experimental results, it has not yet proved possible to calculate a specific value for J in any particular case. The reason for this lies

in the considerable complexity of the general problem of overlapping wave functions in the solid, and we must accept the fact that the coefficients J_{ij} are in reality only empirical parameters.

Moreover we must note that the simple cosine coupling in Eqn (2.1) with a scalar value of J_{ij} rests on theoretical assumptions which are not always justified, and that in particular the value of J will depend on the ionic orbital states if more than one is involved. We shall return to this point again in connection with the problem of magnetic anisotropy in § 3.4, but we want to emphasize already at this stage that Eqn (2.1), which is often taken as an undisputed starting point, is itself an approximation. However for many purposes it provides a very adequate description of the interactions, particularly if we are prepared to treat the J's as adjustable parameters. In this connection it would be interesting to obtain some direct measurements of J's, such as have recently been carried out for antiferromagnetic compounds (Baker and Bleaney 1955, Griffiths *et al.* 1959, Schawlow, Wood and Clogston 1959, Coles, Orton and Owen 1960).

In general the strength of the exchange interaction does not seem to depend very critically on the nature of the anions, provided the atomic spacings are such as to provide a reasonable degree of overlap with the magnetic ions. It is somewhat surprising therefore that practically all the known ferrimagnetic compounds are oxides, though a few sulphur analogues have been prepared (Lotgering 1956 b) and there has recently also been a report of a ferrimagnetic fluoride (Knox and Geller 1958). The reason for this seems to be largely historical and possibly due to the fact that oxides are chemically very stable and easy to handle. On the other hand they are all highly refractory, and this has led to great difficulties in preparing good single crystals, only some of which have now been overcome by using suitable flux techniques (see § 2.3). It may be profitable to stress, therefore, that there is nothing unique about oxides and that some of the problems of sample preparation might well be overcome by trying other types of compounds.

There are three main difficulties in preparing 'adequate' samples for ferrimagnetic investigations. The first is chemical purity in the usual sense. It has been found that relatively small amounts of certain metal impurities can quite seriously affect some properties : thus, for example, 1% cobalt in ferrous ferrite reduces the magnetocrystalline anisotropy by an order of magnitude at room temperature (Bickford, Brownlow and Penoyer 1957), while at very low temperatures even smaller amounts of impurities can have a large effect (see §§ 5.4 and 5.7). Of course once the importance of these factors is recognized it is possible to take special precautions against certain impurities, but it is clear that very careful chemical techniques may be involved. In the past, samples have often been prepared on a rather large scale, using standard ceramic techniques in which accurate control of composition is very difficult, and this has led to much uncertainty in the interpretation of physical results.

The second factor, which is considerably more difficult to guard against, arises from uncertainty in the oxidation state of certain cations. Thus for example it is almost impossible to prepare any compound containing Fe^{3+} and oxygen in such a way as to exclude the possibility of forming a certain amount of Fe^{2+}. The main reason for this lies in the very high temperatures which must be used in order to produce crystalline or highly sintered samples. Much research has gone into

controlling this effect, either by maintaining a suitable atmosphere around the sample during the preparation, or by the use of suitable ' additives ' (e.g. Mn^{2+}) to stabilize the valencies of the ions (see for example Van Uitert 1956). None of these attempts has met with complete success, and it is still necessary to carry out a complete analysis for valency as well as composition to specify any given sample. Although knowledge of the departure from stoichiometry is a useful guide, only complete elimination of this effect will lead to consistently reproducible results.

There are two reasons why departures from an exact valency composition may be serious. If a compound contains the same type of ion in two different valency states it is very easy for two ions to exchange one electron, and electrical conductivity results. This can then give rise to eddy currents and all the unpleasant complications found in metals, and may lead to a spurious frequency dependence of some macroscopic quantities (see for example Bozorth *et al.* 1955). The other, more specialized reason is that certain ions in a particular valency state may have pronounced and undesirable properties, similar to certain chemical impurities. Thus, for example, while Fe^{3+} is loosely coupled to a crystal lattice (being in an S-state), Fe^{2+} is very strongly coupled through crystal field and spin–orbit effects, so that even a small degree of reduction of Fe^{3+} to Fe^{2+} can profoundly influence the total spin–lattice interaction (see § 5.7).

The third main difficulty in preparing suitable samples is that of cation disorder. In certain structures, for example all the spinel type ferrites, one particular crystal site can be occupied by two types of magnetic ion (Fig. 1(*c*)), and conversely one type of ion can, in certain circumstances, distribute itself on two types of crystal sites (Fig. 1(*d*)). The first of these types of disorder is inevitable in this kind of structure and can be avoided only by devising new compounds in which one kind of crystal site accommodates one type of ion exclusively. The second type of disorder may depend on heat treatment of the sample, since even though one type of ion may ' prefer ' a particular crystal site, the activation energy to another site may be small enough to allow some redistribution to take place at high temperatures, and this can be made permanent by sudden cooling or quenching. Thus a wide variety of ion arrangements can be achieved, but reproduced only by very careful control of the heat treatment. This type of disorder is of major importance only in certain compounds (e.g. magnesium ferrite), but it may be that even very small departures from a particular state of ordering can sometimes produce noticeable results. For example in nickel ferrite the Ni^{2+} ions are normally found on one particular crystal site, known as the B or octahedral site. A Ni^{2+} ion on such a site contributes little to the anisotropy of the compound. However, in surroundings of tetrahedral symmetry, corresponding to the other kind of site available to the cations, such an ion would contribute an amount 100 to 1000 times greater, comparable for example with an equivalent amount of a Co^{2+} impurity (Smart 1954, see also § 3.4). Thus even very small departures from the usual arrangement could give rise to considerable discrepancies in the measured anisotropy.

It will be clear from all this that not only is extremely great care needed in preparing samples, but that certain types of substances are more susceptible to particular difficulties than others. It is unfortunate that the spinel type ferrites on which so much effort has been expended during the last 13 years fall into this class, and it seems certain that even though a great deal has been discovered about

ferrimagnetism using ferrites, few quantitative results will remain unaltered. In discussing all experimental results we must therefore consider whether the particular type of measurement is likely to be sensitive to the kinds of ' impurity ' which may have been present.

From a theoretical point of view, too, detailed explanations of phenomena can be attempted only if the experiments relate to well-defined compounds with regular structures. Thus the B site disorder in ferrites entails a gross averaging over two types of ion which can never lead to an exact description of all the physical properties.

We may summarize this section by listing some of the features desirable in ferrimagnetic compounds suitable for detailed physical investigation :

(i) Strong antiferromagnetic coupling between neighbouring, inequivalent magnetic ions.

(ii) Well-defined, regular and preferably simple crystal structure.

(iii) One-to-one relation between the crystal sites and the types of constituent ions.

(iv) Unique valency for each type of ion present ; accurate control of valency in method of preparation.

(v) Simple method of growing large and perfect single crystals.

(vi) Range of compositions possible for the same structure in order to vary magnetic properties.

Although none of the ferrimagnetic compounds known at present fulfils all these requirements, there are good prospects that such substances may soon be synthesized. By far the best of the currently used compounds are the rare earth garnets, which have the disadvantage of a rather complicated structure (see § 2.2) but largely satisfy the other requirements. Their discovery has done much to help the progress in ferrimagnetism.

2.2. *Crystal Structures*

In this section we shall describe briefly the types of structure in which ferrimagnetism has been found and we shall indicate the very large variety of compounds which have been derived from these by substitution. We shall not attempt to describe the details of these structures, some of which are very complex and can best be examined in the original literature, but we shall try to identify those features necessary for an understanding of the magnetic properties.

The most important of these is the type and number of inequivalent cation sites in a unit cell, since this often provides a useful guide to the number of magnetic sub-lattices in the ferrimagnetic state. The point symmetry at each cation site is also of considerable importance since the magnetic properties of certain ions are completely altered by a change in local symmetry. This kind of consideration properly belongs to the field of paramagnetism and reference may be made to reviews by Van Vleck (1932), Bleaney and Stevens (1953) and Bowers and Owen (1955). The effects of the local symmetries are combined by the overall symmetry of the compound, and this controls the form and magnitude of the macroscopic magnetic properties, such as the crystalline anisotropy (see § 3.4).

A striking converse of the effect of local symmetry on the magnetic properties occurs in some compounds in which the presence of certain magnetic ions causes a

deformation of the whole crystal structure through their interaction with their local environment. This sort of effect is closely bound up with the very complex problem of chemical structure, and reference may be made to extensive discussions by van Santen and van Wieringen (1952), Romeijn (1953), Goodenough and Loeb (1955), Dunitz and Orgel (1957), McClure (1957), Wickham and Croft (1958), Wojtowicz (1959), Kanamori (1960) and others referred to in their papers. We shall later refer to these publications collectively as I. From our point of view we must bear in mind that in compounds which show unusually large local distortions the properties of the magnetic ions may also be considerably altered, and care must then be taken in interpreting magnetic data. One example of this is 'spin quenching' proposed by Baltzer and Wojtowicz (1959), which is discussed in § 3.2. In considering structures of magnetic compounds below their Néel point one must always bear in mind that the magnetization vector provides a unique direction, which through magnetostriction can give rise to local distortion of the structure.

Another property closely related to the effect of local environment is the preference of certain ions for particular crystal sites, which is important in determining the ionic arrangement in mixed crystals, as well as the ease of migration of ions at high temperature. This problem is also discussed in the above references, as well as by Verwey and his co-workers (Verwey, Haayman and Romeijn 1947, Verwey and Heilman 1947), and in most cases a reasonable explanation for particular preferences can be deduced. From a practical point of view the most direct evidence can often be inferred from magnetic measurements, though the results of neutron diffraction can be even more definite (see § 6.1).

One structural feature which has often been discussed in connection with ferrimagnetic compounds is the number and kinds of metal–anion linkages. Detailed knowledge of these is important for an understanding of particular types of indirect exchange mechanism, but until there is some direct experimental information or the theories are made more quantitative an analysis of this kind does not seem worth while.

We shall now consider in turn the various structural types in which ferrimagnetism has been found up to the present.

(*a*) *Spinel.* This is the structure of the ferrites which comprise the majority of all known ferrimagnetic compounds. The general formula is PQ_2X_4, where $X = O^{2-}$, S^{2-} or Se^{2-} and PQ_2 can be a wide variety of combinations of metal ions of various valencies. A list of these has been given by Gorter (1954), together with detailed drawings of the structure and earlier references. In the most common ferrites P is a divalent ion such as Mn^{2+}, Fe^{2+}, Ni^{2+}, Co^{2+}, Cu^{2+}, Zn^{2+} or Mg^{2+} and Q is a trivalent ion such as Mn^{3+}, Fe^{3+}, Co^{3+}, Al^{3+} or Ga^{3+}. Mixed crystals with these ions can also be prepared. At high temperatures the structure of all the spinels is cubic with eight PQ_2X_4 molecules in one unit cell, which belongs to the space group O_h^7. The length of the cell edge is about 8Å. The structure can be described as a slightly distorted, close packed hexagonal arrangement of the X ions with the metal ions filling half the total number of interstices. The occupied cation sites fall into two classes, eight having four nearest anion neighbours giving approximately tetrahedral cubic symmetry, and sixteen sites having six anion neighbours, giving approximately octahedral cubic symmetry. These sites are commonly called the A and B sites. It might be assumed from a cursory inspection

of the chemical formula and the number of anion sites that the P type ions would occupy the A sites and the Q type ions the B sites. However, Barth and Posnjak (1932) pointed out that such a ' normal ' arrangement in fact occurs only in some spinels, and that in others half the B sites are filled by P ions and half by Q ions in a random way, while the remaining Q ions occupy the A sites. Such an arrangement is termed ' inverse ', and in certain cases an intermediate arrangement, or partial inversion, is possible. From our point of view this difference is very important since with a few exceptions (e.g. $CuCr_2O_4$, Mn_3O_4) only inverse, or partly inverse, spinels are ferrimagnetic. Details of the different cation preferences are discussed in the references (I) cited above.

In terms of crystal field theory the difference between the A and B sites can be described by a change of sign of the cubic electrostatic potential, which determines the sequence of orbital energy levels. This is of considerable importance in connection with magnetic moments, susceptibilities and contributions to anisotropy and will be discussed further in §§ 3.1, 3.2 and 3.4. For some properties the main effects stem from the small deviations from perfectly cubic point symmetry at the A and B sites. These can arise in several different ways :

In most spinels there is a small deviation from perfect close packing of the anions. This can be described by a single parameter u, whose difference from the ideal value 0·375 gives a measure of the deviation. Determination of this parameter is difficult, but accurate values can be obtained by combining x-ray and neutron diffraction measurements (see § 6.1). The other departures from cubic point symmetry arise from the metal ions. Due to the fact that not all the interstices in the anion lattice are occupied there is a difference in the *cation* environment around any cation site, and to a first approximation this can be described as a trigonal distortion along one of the four [111] axes. In a ' normal ' spinel this is the only effect, but in the inverse structure the random distribution of P and Q ions on the B sites provides a range of such local distortions. The importance of these distortions is hard to assess, as the usual methods of structure determination provide only a statistical average for small variations. An alternative approach might be to investigate the paramagnetic resonance spectra of tracer amounts of magnetic impurities in inverse diamagnetic spinels. In any case the random distribution provides a very unpleasant complication in the detailed explanation of certain magnetic bulk properties.

At low temperatures some spinels show a departure from overall cubic symmetry, and we may distinguish two main types, in addition to possible effects from magnetostriction. The most pronounced of these occurs in compounds containing ions with an orbital degeneracy, the removal of which can provide the strain energy needed to deform the whole crystal. The deformation is usually to tetragonal symmetry with the axis parallel to one of the original cube axes, and c/a ratios of the order of 0·9 to 1·1 are found. Examples of compounds of this kind are $CuFe_2O_4$ (Weil, Bertaut and Bochirol 1950, Prince and Treuting 1956), $CuCr_2O_4$ (Bertaut and Delorme 1954, Prince 1957 a) and Mn_3O_4 (Mason 1947, Kasper 1959). For a discussion of the effect and further examples see I.

The second type of mechanism responsible for the distortions occurs in Fe_3O_4 near 120°K, and this may also cause distortions in other spinels containing both Fe^{2+} and Fe^{3+} ions. In such substances there is, at high temperatures, a rapid

interchange of electrons between the Fe ions, as indicated by a relatively high electrical conductivity, and we cannot identify any Fe^{2+} or Fe^{3+} ions on specific crystal sites. Below a certain temperature however this electron transfer ceases, with a corresponding abrupt drop in conductivity, and long-range ordering of the two types of Fe ions sets in. In Fe_3O_4 this leads to an orthorhombic structure for the whole crystal, with an axis along the old cubic [110] direction, but this is complicated by the fact that six of these axes are equally likely, so that twinning can occur. The twinning may be suppressed by cooling through the transition in the presence of a suitably oriented magnetic field. This complex phenomenon has been extensively investigated by Verwey, Haayman and Romeijn (1947), Bickford (1950), Williams, Bozorth and Goertz (1953), Calhoun (1954), Hamilton (1958) and others.

(*b*) *Garnets.* The structure of these compounds is rather complex with eight molecules of $P_3Q_2R_3O_{12}$ in a unit cell, and it is really best discussed using a three-dimensional model. However it has a very high degree of symmetry and none of the randomness of the inverse spinels. The unit cell is cubic with space group O_h^{10}–Ia3d, and it can be considered in terms of eight octants which are related to one another by simple 180° rotations. Each octant has a three-fold rotation axis along one of its body diagonals and the whole structure has a centre of inversion. The metal ions P, Q and R are located in sites which have respectively eight, six and four nearest oxygen neighbours and are known as the 24c, 16a and 24d sites. It is often convenient to denote the types of site in a formula by the bracket notation : 24c = { } ; 16a = [ ] ; 24d = ( ), i.e. $\{P_3\}[Q_2](R_3)O_{12}$ (Gilleo and Geller 1959). None of the oxygen polyhedra around the cations is regular but they approximate fairly closely to regular cubes, octahedra and tetrahedra. More precisely the point symmetries are D_2 for 24c, C_{3i} for 16a and S_4 for 24d, with the symmetry axes of 16a and 24d respectively parallel to [111] and [001], and the twofold axes of 24c parallel to [001] and [110], together with permutations of the indices implied by the overall symmetry. (All the polyhedra of one type are identical apart from orientation.)

The combinations of P, Q and R which are possible within this structure are very large in number and seem to be limited mainly by ionic size and *total* charge balance within the whole structural formula. There is no simple relation between the coordination of a particular site and the valency of the cations which it will accommodate. Thus, in the mineral garnets found in nature (Menzer 1929) P is a divalent ion such as Ca^{2+}, Mn^{2+} or Fe^{2+}, Q is a trivalent ion such as Fe^{3+}, Cr^{3+} or Al^{3+}, and R is tetravalent Si^{4+}. Yoder and Keith (1951) discovered that it was possible to replace the divalent and tetravalent ions simultaneously by two trivalent ions, and they synthesized $Y_3Al_2Al_3O_{12}$ with the same structure as $Mn_3Al_2Si_3O_{12}$. Independently of this work, and independently of one another, Bertaut and Forrat (1956 a) and Geller and Gilleo (1957 a) succeeded in identifying the magnetic analogue of this compound $Y_3Fe_2Fe_3O_{12}$, which is now usually known as YIG (yttrium iron garnet), and from this derived the rare earth iron garnets (RIG's). Not all the rare earths form garnets with iron, and only those from Sm to Lu have been prepared in a pure state. However it seems possible to include some of the larger ions in the structure provided there are also some smaller ions such as Y. Thus Bertaut and Forrat have reported $Y_{1.5}Nd_{1.5}Fe_2Fe_3O_{12}$.

Also if the Fe^{3+} is replaced by Al^{3+} or Ga^{3+} some larger ions can be included. Lattice parameters of several of these structures have been given by Bertaut and Forrat (1956 a, 1957).

Another simple substitution which can be made in the garnet structure is to replace Si^{4+} with Ge^{4+} to make compounds of the type $M^{2+}{}_3M^{3+}{}_2Ge_3O_{12}$, with M^{2+} = Ca or Mn, and M^{3+} = Fe, Al, Cr as in the natural garnets (Tauber, Banks and Kedesy 1958, Geller, Miller and Treuting 1960). However as in the case of $Mn_3Fe_2Si_3O_{12}$ none of these compounds shows any appreciable ferrimagnetism above 1·3°K (Bozorth and Geller 1959).

Since the first discovery of YIG there have been several improvements in the determination of the structure. In the $O_h{}^{10}$ space group the metal ions are all at non-parameter positions and are specified by the single cubic lattice constant a_0 which is about 12·4Å. This parameter varies with temperature and a thermal expansion coefficient about 10^{-5} deg^{-1} has been reported by Geller and Gilleo (1957 b). The positions of the oxygen ions depend, in addition, on three further parameters (x, y and z). Details of the determination of x, y and z (using x-rays and neutrons) have been discussed by Bertaut *et al.* (1956) and by Geller and Gilleo (1957 b), who also quote some unpublished results of Prince. Using the values x, y and z Geller and Gilleo (1957 b) calculated the exact shapes of the O^{2-} polyhedra and give useful drawings and tables of interionic distances.

All the compounds mentioned above have the great advantage, from a theoretical point of view, that each lattice site accommodates one type of ion exclusively. More recently many new types of garnet compounds have been reported in which this unique occupation scheme does not hold, thereby making a large number of combinations of ingredients possible. An excellent review of these, together with a comprehensive list of references, has been given by Geller (1960). These compounds are of interest in connection with their structures and cation distributions, and also the insight they afford into various types of magnetic ion interactions. However from our point of view most of this work is still too preliminary to be of direct application to the basic problems of ferrimagnetism.

(*c*) *Magnetoplumbite.* This is the structure of compounds of the type $AO.6B_2O_3$ where A = Ba^{2+}, Sr^{2+} or Pb^{2+} and B = Fe^{3+}, Al^{3+}, Ga^{3+} or Cr^{3+}. The composition $BaO.6Fe_2O_3$ is commercially known as ' Ferroxdure ', sometimes referred to as the hexagonal M compound (see below). It is a somewhat complicated hexagonal structure with space group $D_{6h}{}^4$–C6/mmc, which was first investigated by Adelsköld (1938). More recently it has been examined by Bertaut, Deschamps and Pauthenet (1959) and by Went *et al.* (1952) who give some useful explanatory drawings. The unit cell contains two formula units and the twenty-four magnetic B type ions are distributed over five different types of sites. Of these the $4f_1$ are tetrahedral, the 2a, $4f_2$ and 12k are octahedral, while the 2b sites are hexahedral with five nearest oxygen neighbours in the form of a simple bipyramid. The importance of this unusual coordination on the magnetic properties has been discussed by Smit (1959). When B represents a single element there is no randomness in the structure. The site preferences for various ions in mixed crystals have been discussed by Bertaut *et al.* (1959).

The importance of this type of structure is that it has a unique axis, which allows the magnetic anisotropy to be much larger than in overall cubic compounds,

in which it enters only as a second-order effect. These materials are therefore all magnetically ' hard ' and are used, for example, for making permanent magnets.

(*d*) *Complex hexagonal oxides.* Extensive investigations by Braun (1957) have shown that there exist several other types of hexagonal oxides, whose structures are related to combinations of spinel and magnetoplumbite type blocks. Three new series of compounds were recognized and termed W, Y and Z. The compositions of these are rather complicated and can best be understood in terms of a ternary phase diagram of BaO, Fe_2O_3 and MeO, where Me is one of the divalent metals which form spinels with iron oxide. This is shown in Fig. 3. The point marked S

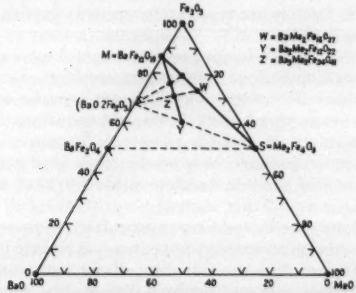


Fig. 3. Ternary phase diagram showing the relation between the hexagonal oxides M, W, Y, and Z, their constituent oxides and the spinel phase S (*after* Smit and Wijn 1959).

(1)	(2)	(3)	(4)	(5)	(6)	(7)
M	$BaFe_{12}O_{19}$ or $BaO.6Fe_2O_3$		2	2	9	1
W	$BaMe_2Fe_{16}O_{27}$ or $BaO.2MeO.8Fe_2O_3$	W = M + 2S	2	4	13	1
S	$MeFe_2O_4$ or $MeO.Fe_2O_3$		8	1	2	0
Z	$Ba_3Me_2Fe_{24}O_{41}$ or $3BaO.2MeO.12Fe_2O_3$	Z = M + Y	2	6	19	1
Y	$Ba_2Me_2Fe_{12}O_{22}$ or $2BaO.2MeO.6Fe_2O_3$		3	4	10	0

(1) Compound ; (2) chemical formula ; (3) chemical inter-relation ; (4) number of chemical formulae per unit cell ; (5), (6), (7) number of occupied sites per unit cell : (5) tetrahedral, (6) octahedral, (7) hexahedral (5 O^{2-} neighbours).

Me = Mn, Fe, Co, Ni, Zn, Mg or mixtures of these.

represents the spinel composition and M the magnetoplumbite structure. The chemical formulae of the compounds are given in the Table, which also shows some of the chemical inter-relations and numbers of lattice sites for the magnetic ions. The crystallography of these compounds is rather complex and reference must be made to papers by Braun (1957) and Jonker, Wijn and Braun (1956, 1957) for details and explanatory drawings. Several of these compounds are of considerable interest in connection with their anisotropy which is both large and unusual in sign. This will be discussed further in § 3.4.

(*e*) *Ilmenites.* Although the natural mineral ilmenite, $FeTiO_3$, is antiferromagnetic, two ferrimagnetic compounds have recently been reported which have

the same structure. They are $NiMnO_3$ and $CoMnO_3$ (Cloud 1958, Bertaut and Forrat 1958, Swoboda, Vaughan and Toole 1958). The structure is very simple, with a space group C_{3i}^2–$R\bar{3}$. It belongs to the rhombohedral class and can be described by chains of alternate Ni and Mn ions, each surrounded by six oxygen ions in an almost regular octahedron. There is some disagreement regarding the exact positions of the anions, but there is general agreement over the overall symmetry, which is unusually high. In fact from several points of view this is one of the simplest ferrimagnetic structures and it would be interesting to see if other isomorphous compounds can be prepared. However even the preparation of these two compounds is subject to considerable difficulties (see § 2.3).

(*f*) *Fluorides.* So far only one type of ferrimagnetic fluoride has been reported : $Na_5Fe_3F_{14}$ (Knox and Geller 1958). The structure has two forms, a relatively simple high temperature form belonging to the tetragonal space group $D_4(6)$–$P4_22_12$, and a very complicated monoclinic low temperature form, whose space group has not yet been identified definitely. The high temperature form, which can be obtained by rapidly quenching a melt, has a unit cell which contains $6Fe^{3+}$ ions in two inequivalent sites : 2a and 4d. The coordination of both these sites is octahedral. The low temperature form on the other hand has $72Fe^{3+}$ ions on at least nine inequivalent sites and it is therefore unlikely to be of much further fundamental interest. Knox and Geller showed also that the Fe^{3+} could be replaced by Co^{3+} or Cr^{3+}, but only the Fe and Co compounds were ferrimagnetic above 20°K. Further work on these very interesting compounds was reported to be in progress.

(*g*) *Pyrrhotite,* Fe_7S_8. This is one phase of the very complex iron–sulphur system, whose magnetic properties have been discussed by Néel (1953), Lotgering (1956 a) and Hirahara and Murakami (1958) who give references to earlier work. This system which otherwise has many unpleasant features, such as thermal hysteresis, is interesting because of the unusual origin of the ferrimagnetic moment. From a detailed investigation of the crystal structure Bertaut (1952) deduced that the lattice contains cation vacancies which are ordered, to give a complex structure with a monoclinic unit cell containing eight molecules of Fe_7S_8. Néel (1953) pointed out that this ordering corresponds to vacancies occurring on only one of two otherwise equivalent antiferromagnetic sub-lattices, suggesting that the vacancies are responsible for the observed magnetic moment. However the precise role of vacancies is inevitably very complicated as the question of charge compensation and local valency must be considered. Further information on the properties of this complex system has recently been obtained by Sparks *et al.* (1960) who used neutron diffraction to study the spin alignment.

2.3. *Methods of Preparation*

There are at least as many methods of preparing ferrimagnetic compounds as there are laboratories working in this field. This applies particularly to the preparation of polycrystalline samples on which by far the most work has been done, and many ingenious variations have been devised for special technical applications. However for basic physical investigations pure single phase compounds are required and these are very difficult to obtain, as even for a well established process extremely stringent control of the exact conditions is essential to ensure reproducible results. For any particular experiment it is essential

therefore to devise a set of analytical criteria to which each specimen must be subjected. Some of the most common tests of 'quality' include: (i) Chemical analysis for (*a*) cation composition, (*b*) cation–anion ratio (average valency), (*c*) chemical impurities. (ii) Crystallographic investigation: (*a*) check of lattice parameters, (*b*) absence of foreign phases by x-rays, etch figures and optical examination, (*c*) density. (iii) Electrical conductivity (a.c. and d.c.) to indicate departures from stoichiometry. (iv) Check of known magnetic properties (magnetic moment, resonance line width, etc.).

2.3.1. *Polycrystals.*

All methods of making polycrystalline samples are basically very similar and depend on well-known ceramic techniques.† The starting ingredients, usually in the form of oxides, are first intimately mixed, either by ball milling, or by chemical solution followed by simultaneous co-precipitation. The mixture is then calcined for several hours at between 700° and 1100°c, and a solid–solid reaction forms the mixed compound. Sometimes this process is repeated several times until there is evidence of complete reaction. At this stage the product is a finely divided powder and for most purposes it is then necessary to compress and refire it at a higher temperature (1100–1400°c) to obtain a dense sintered solid. Among the many factors which must be carefully controlled are the firing temperatures and times, cooling rates, atmosphere in the furnace to prevent oxidation or reduction, avoidance of impurities during the mixing processes (pickup from ball mills), particle size distribution before the final pressing (to obtain a high density), accurate control of purity and composition of starting materials (errors in weighing due to adsorbed water or CO_2). These and other important factors have been discussed by Snoek (1947 a), Fresh (1956), Van Uitert (1956), Pippin and Hogan (1958), Wolf and Rodrigue (1958) and Economos (1959), who also give earlier references. At present the exact procedure for any particular compound can only be found by trial and error, followed by analysis of the product.

2.3.2. *Single crystals.*

For some investigations single crystals are essential, and for many they are desirable both to avoid the effects of anisotropy and to obtain higher purity. During recent years several different methods have been tried, but although these have met with some success the standard of ferrimagnetic single crystals is still very far below that which has been achieved for other types of crystals. As mentioned before this is partly due to the limited range of known ferrimagnetic compounds. There are basically two types of methods which have been used, growing from the melt and growing from solution, and for each of these there are several variants.

(*a*) *The Bridgman–Stockbarger method.* This is the well-known method used for growing large single crystals of low melting point salts, such as LiF. The compound is heated to above its melting point in a crucible, the lower end of which is pointed, and the crucible is then lowered slowly to a cooler part of the furnace.

† *Note added in proof*: A novel method involving the oxidation of suitable alloys at the anode of an electrolytic cell has recently been reported by Forestier and Petit (1960). Only brief details regarding purity, etc., have been given. Another interesting, but not very reproducible method which has been reported (Wencus and Leavitt 1956) consists in 'flame-spraying' an alcohol solution of mixed metal nitrates, followed by sintering of the very fine powder produced.

Nucleation of seeds first takes place at the point and usually one of these crystals then develops to the exclusion of all others. In this way extremely large and perfect crystals can be grown, provided the substance is stable near its melting point and it also melts congruently. Unfortunately most oxide materials have very high melting points, at which the equilibrium pressure of oxygen is so high that reduction of some of the ions inevitably occurs. As we have seen this can produce undesirable electrical properties, and in any case it will prevent a pure single phase from developing. An exception to these considerations is the compound Fe_3O_4, magnetite, which has a relatively low melting point 1575°c, and already contains such a large proportion of divalent iron that it is stable at this temperature. Large crystals of this substance have been grown by Smiltens (1952). In order to grow other oxides containing Fe^{3+} but free from Fe^{2+} very high oxygen pressures would be needed and these would be difficult to maintain at high temperatures. A possible solution of this problem has been suggested by Wijn (1952) who grew crystals of $BaFe_{12}O_{19}$, $MnFe_2O_4$ and $BaFe_{18}O_{27}$ from the melt, using induction heating. He pointed out that with this method of heating it might be easier to maintain adequate oxygen pressures, but so far there have been no reports of any such experiments.

(*b*) *Verneuil flame fusion method.* A method of melting more refractory solids is to use a high temperature flame (Verneuil 1904). The ingredients are mixed and finely powdered and slowly fed into the gas stream of an oxy-hydrogen flame, which is directed towards the end of a small pillar of refractory material. The powder melts in the flame and begins to collect on the pillar, which is then slowly withdrawn from the flame. As part of the melt moves to a cooler region of the flame it crystallizes, and as more material is added an equilibrium state can be set up in which crystallization proceeds continuously. In order to grow single crystals by this method very careful control of the gas flow, solid feed and withdrawal rate are required, but for many substances the necessary conditions are now well known. Unfortunately in applying this method to the known ferrimagnetic substances we are again faced with the problem of cation reduction and equilibrium oxygen pressures, over which we now have even less control than in the Bridgman method. Thus although it has been possible to grow large single crystals of ferrites by this method, it is found that they are usually far from stoichiometric, containing large amounts of Fe^{2+}. Nevertheless this method has provided the source of many of the single crystals on which physical measurements have been reported in the literature, and in the absence of a definite valency analysis all such results must be considered suspect.

Another problem which can arise in all methods of growing from the melt is that the substance melts incongruently, i.e. separates into two liquid phases having different compositions. Rudness and Kebler (1960) have recently shown that this difficulty can be overcome in the flame fusion method by suitably adjusting the feed conditions, and they have succeeded in growing large and almost stoichiometric crystals of yttrium iron garnet.

(*c*) *Flux methods.* A common and simple method of growing crystals is from solution, either by evaporation or slow cooling, but for oxide materials this requires some unusual solvents. For example, Remeika (1956) found empirically that molten PbO will readily dissolve most transition group and rare earth oxides, and

he succeeded in growing ferrite crystals by slow cooling of mixtures of divalent oxides, Fe_2O_3 and PbO. These crystals are generally much smaller than those grown from the melt, but they are considerably purer from a valency point of view. They do contain small impurities of lead from the flux and platinum from the crucible, but these do not greatly affect most magnetic properties. The crystallization process has since been studied extensively by Nielsen and Dearborn (1958), who investigated the ternary phase diagram of PbO–Y_2O_3–Fe_2O_3 with a view to growing single crystals of YIG by this method. They showed that the mechanism is considerably more complicated than a simple deposition from solution, since other phases of the ternary system are also crystallized. In order to obtain good crystals it is necessary to use a very steady and slow cooling rate (2 to 5 deg hr^{-1}) over a temperature range from 1325 to 900°c, so that a single run may require several weeks. Further experimental details are given by Nielsen (1958) and Nielsen and Dearborn (1958). An important improvement of the method has recently been reported by Nielsen (1960) who found that the addition of PbF_2 to the PbO considerably increases the size of the crystals formed by decreasing their number, presumably through a reduction of the viscosity of the melt.

It is clear that there is nothing unique about PbO as a flux and several workers have reported success in growing ferrite crystals using other 'solvents'. These include borax ($Na_2B_4O_7$) (Galt, Matthias and Remeika 1950, Okamura and Kojima 1952) and Na_2CO_3 (Mones and Banks 1958), but the crystallization process for these substances is not clear. Undoubtedly other and better fluxes will be discovered by further research. So far all experiments with fluxes have been based on the slow cooling method of crystallization, but there seems to be no reason against using controlled evaporation of the solvent at constant temperature, as for aqueous solutions.

(*d*) *Hydrothermal methods.* Under conditions of high temperatures (~700°c) and very high pressures (~3000 atm) even oxides will dissolve in water and crystals can be grown. So far this method has been little used for magnetic materials, but Swoboda, Vaughan and Toole (1958) have reported some success in growing the ilmenites $NiMnO_3$ and $CoMnO_3$, which have not been prepared by any other method. The combination of high pressures and temperatures requires some special precautions, but otherwise this method would seem to have considerable potential advantages over the others for preparing very pure samples.

§3. Magnetostatic Properties

3.1. *High Temperature Susceptibility*

The first important success of Néel's theory was the explanation of the anomalous susceptibility–temperature curves which had been observed at high temperatures for substances known to become spontaneously magnetized at lower temperatures. In contrast to the normal ferromagnetic metals, which obey a Curie–Weiss law of the form $\chi = C/T - \theta$, with $\theta > 0$, several ferrites had been observed (Serres 1932) to follow this law with a *negative* value of θ at very high temperatures, but marked deviations towards greater susceptibilities at lower temperatures (see Fig. 4). From the asymptotic behaviour of this pseudo-hyperbolic variation, Néel concluded that the principal interactions between the magnetic ions are antiferromagnetic,

and he showed how such a coupling could lead to the observed curves. The mathematical details of Néel's theory both for the susceptibility and magnetization have been reproduced frequently (see for example Fairweather, Welch and Roberts 1952, Smit and Wijn 1959) and we shall therefore mention here only its outline, and examine some of the assumptions and possible improvements.

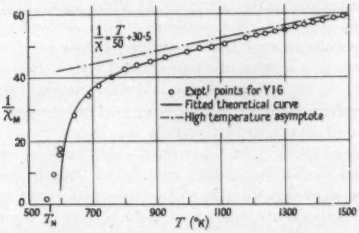


Fig. 4. Typical susceptibility–temperature curve for a ferrimagnetic material above its Néel point (*after* Aléonard and Barbier 1959).

Néel's theory starts from the molecular field approximation that the effect of the exchange interaction on any particular ion can be replaced by the sum of effective magnetic fields, whose magnitudes are proportional to the average magnetic moments of the neighbouring ions. By grouping the neighbours as belonging to one of two types A and B these can be expressed in terms of three molecular field constants: $n_{AB} = n$; $n_{AA} = \alpha n$; $n_{BB} = \beta n$, to be determined by experiment. A simple calculation of the susceptibility then shows that $1/\chi$ should follow a law of the form

$$\frac{1}{\chi} = \frac{T}{C} + \frac{1}{\chi_0} - \frac{\sigma}{T-\theta} \qquad \ldots\ldots(3.1)$$

where χ_0, σ and θ are functions of n, α, β, and the fractions of ions belonging to groups A and B. The constant C is the average paramagnetic Curie constant of the ions. If all orbital angular momentum is quenched, as is common in many iron group salts,

$$C = \sum_i \frac{N_i g_i^2 \mu_B^2 S_i(S_i+1)}{3k}, \qquad \ldots\ldots(3.2)$$

where N_i is the number of ions having spin S_i, μ_B is the Bohr magneton, k Boltzmann's constant and g_i the spectroscopic splitting factor which is almost exactly equal to 2. An extension of the theory to the case of three coupled sublattices has been given by Aléonard and Barbier (1957, 1959).

Equation (3.1) can be fitted quite closely to the experimental curves (Néel 1948 a, Foëx and Serres 1950, Fallot and Maroni 1951, McGuire, Howard and Smart 1952, Aléonard, Barbier and Pauthenet 1956, Borovic-Romanov and Orlova 1957, Aléonard 1960), but such a fit is not very sensitive to the exact values of n, α and β,

and some caution should be exercised in using the values so deduced in other calculations. In the absence of any detailed information on the exchange interactions the only direct comparison which can be made between the simple theory and experiment is for the Curie constant C, and for this quite appreciable discrepancies are generally found (see for example Foëx and Serres 1950). One of the main reasons for these was explained by Néel (1949, 1951) who pointed out that thermal expansion will affect the strength of the exchange interaction, so that the molecular field constants will vary with temperature. He made the simple assumption that all the molecular field coefficients vary linearly with temperature at the same rate, so that α and β are constant and $n_T = n_0(1+\gamma T)$, where γ is a negative constant of the order of magnitude of the thermal expansion coefficient. With this assumption the apparent Curie constant C' is related to the true paramagnetic C by the relation

$$\frac{1}{C'} = \frac{1}{C} + \frac{\gamma}{\chi_0},$$

and the discrepancy with the experimental value can be removed by choosing a suitable value for γ. This procedure leads to values of γ which are qualitatively quite reasonable. Thus for yttrium iron garnet γ is found to be -1×10^{-4} (Aléonard and Barbier 1957) which is about ten times the linear expansion coefficient found by Geller and Gilleo (1957 b), implying that the superexchange interaction varies roughly as the tenth power of the distance between the ions. In addition to this effect there are at least four other reasons why the simple theory for the susceptibility should be modified in certain cases.

In Eqn (3.2) it is assumed that all orbital angular momentum is quenched, but this is not always strictly true. Incomplete quenching of the orbital angular momentum leads to changes in the value of g from 2, as is well known in paramagnetism, and so changes C. Another effect arises when the energy levels of some of the ions are not simply $(2S+1)$-fold degenerate (for example in Co^{2+}), when thermal repopulation gives rise to deviations from Curie's law even in the absence of interactions. The effect of excited energy levels can also enter in another way, which so far seems to have been ignored. At high temperatures the measured susceptibilities become quite small and there may be appreciable temperature independent contributions from states at quite high energies (Van Vleck 1932). For example in all the rare earth ions (except Gd^{3+}) there are low lying states of **J** which are at energies Δk of the order of 10^3 to $10^4 k$, and these will give contributions to χ of the order of C/Δ, compared with the Curie law value C/T. The effect will generally be smaller for transition group ions, but for some (e.g. Cu^{2+}) it could be quite large, and it may be comparable with the effect of interactions. A similar, rather smaller correction arises from diamagnetism.

A completely different effect arises from the inadequacy of the molecular field theory, especially near the Néel point, but so far only two attempts have been made to improve the statistical calculations. The first was made by Smart (1956) who adapted to the case of ferrimagnetism the Bethe–Peierls cluster method, as applied by Weiss (1948) to ferromagnetism and Li (1951) to antiferromagnetism. Making considerable simplifying assumptions he compared the two methods for a hypothetical example and showed that they differ appreciably even at temperatures

as high as $4T_N$. Thus for the normal ferrimagnetic compounds, which have Néel points above room temperature, all the susceptibility measurements are made in a region where the molecular field theory appears to be not very accurate. Smart's calculation may have been somewhat pessimistic as he considered the case of $S = \frac{1}{2}$, for which one could expect quantum mechanical effects to be more pronounced than for larger values of the spin, and it would be interesting to repeat the calculations for more realistic cases (cf. similar calculations for antiferromagnetic systems, Smart 1959). On the other hand Smart's simple treatment shows quite clearly that there will be some appreciable deviations from the simple theory, and that some refinements are essential before a detailed analysis of the results is meaningful. More recently, Wojtowicz (1960) has derived a series expansion for the susceptibility in powers of $1/T$ up to $1/T^5$. His results are qualitatively similar to Smart's and again emphasize the inadequacy of the molecular field treatment, but it will be difficult and tedious to fit the complex analytic expressions which he obtained to empirical data.

The present position therefore is that the simple Néel molecular field theory gives an adequate qualitative explanation of the shape of the susceptibility curves, which can be fitted by choosing a sufficient number of strictly empirical constants, but for a detailed comparison there is still no simple method.

3.2. *Spontaneous Magnetization*

The second important success of the Néel theory was the quantitative explanation of the magnitude of the spontaneous magnetic moments observed in many ferrite type materials, and their variation with temperature. We shall consider these two aspects of the theory separately and we shall discuss first the low temperature limit of the moment, extrapolated to absolute zero. In connection with this extrapolation we should add a word of caution : in some early references the extrapolation was made from measurements at relatively high temperatures and in a manner contrary to the thermodynamic condition $\partial M_s/\partial T = 0$ at $T = 0°\text{K}$, so that some values in the literature are therefore high. More recently measurements have usually been carried down to much lower temperatures, but then trouble is sometimes experienced from anisotropy which becomes large and makes it difficult to saturate the sample. This can be overcome by using single crystals and larger measuring fields. The uncertainty can also be reduced by employing more reliable extrapolation formulae over a wide range of temperatures. With care, any discrepancy between experimental and theoretical values can now usually be ascribed entirely to the theory.

3.2.1. *The magnetic moment at* $T = 0°\text{K}$.

The magnetic moment in the completely ordered states shown in Figs 1(*a*), (*b*) and (*c*) follows directly from a knowledge of the atomic magnetic moments of the magnetic ions. In the 3d transition group these are often close to the ' spin-only ' value $2\mu_B S$, and the moment can be calculated from the number of ions on inequivalent sites. Thus in yttrium iron garnet, in which there are three Fe^{3+} ions on d sites for every two on a sites, one would expect (Aléonard, Barbier and Pauthenet 1956) a magnetic moment of $(3-2)\frac{5}{2}2\mu_B = 5\mu_B$, for every formula

unit of $Y_3Fe_5O_{12}$. This may be compared with the best experimental value (Gilleo and Geller 1958) of $4{\cdot}96\mu_B$. However such excellent agreement is rare, and although the simple theory generally accounts well for many observed moments, discrepancies of the order of 10% are common. In a few cases the discrepancies are much bigger. There are a large number of different factors which may account for these in various circumstances.

As in the case of the susceptibility, small admixtures of excited orbital states through the spin–orbit coupling can cause the g-value to differ from two, and this changes the effective magnetic moment to $g\mu_B S$. This effect almost certainly explains the observed moment of nickel ferrite $(Fe^{3+})\,[Ni^{2+}Fe^{3+}]O_4$, which is $2{\cdot}2\mu_B$ compared with the spin-only value of $[2\mu_B\uparrow + 5\mu_B\uparrow] - (5\mu_B\downarrow) = 2\mu_B\uparrow$ (Pauthenet 1952). Similarly the magnetic moment of the Co^{2+} ions in cobalt ferrite exceeds the spin-only value of $3\mu_B$ by about $0{\cdot}7\mu_B$, but in this case the effect is complicated by low lying orbital energy levels which depend on local distortions from cubic symmetry. Considering the ideal case of small amounts of Co^{2+} in magnetite, Slonczewski (1958) showed that this will lead to an orbital contribution to the total magnetic moment which has the right order of magnitude, but which is anisotropic even though the crystal as a whole has cubic symmetry. (This effect cannot arise in the susceptibility which must be isotropic under cubic symmetry.) For the case of pure cobalt ferrite Slonczewski's model will be complicated by the B site disorder in the inverse spinel structure, but a somewhat anisotropic moment would still be expected. So far this has not been observed. A more extreme case of orbital effects on the magnetic moment occurs in $CoMnO_3$ for which antiparallel alignment of Co and Mn leads to zero magnetic moment in the ' spin-only ' approximation. The observed moment of $0{\cdot}72\mu_B$ is therefore ascribed entirely to orbital angular momentum (Bozorth and Walsh 1958).

Another class of compounds in which there are considerable contributions from the orbital momentum, modified by interaction with the crystal lattice, are the rare earth iron garnets. Pauthenet (1958) has shown that at temperatures down to about 50°K the rare earth ions behave very much like free particles with moments $g_J\mathbf{J}\mu_B$, where g_J is the Landé factor and $\mathbf{J}$, the total angular momentum, is equal to $\mathbf{L}+\mathbf{S}$. These are loosely coupled antiparallel to the resultant of the two Fe^{3+} sub-lattices, which are practically unaffected and behave as in YIG. At $T = 0$°K one might therefore expect a moment $[3g_J J - (3-2)5]\,\mu_B$† per $R_3Fe_5O_{12}$ molecule. However, experimentally this is only found for $R = Gd^{3+}$, for which $L = 0$ and $J = S$, and for all other ions there are large differences due to the partial quenching of J by the crystalline field. The most extreme case is $Yb_3Fe_5O_{12}$ for which an unquenched J would lead to $[3\tfrac{8}{2}\tfrac{7}{2} - (3-2)5]\,\mu_B = 7\mu_B$, compared with the observed value of $0{\cdot}25\mu_B$ (Pauthenet 1958). Ayant and Thomas (1959 a, 1960) and Wolf (1959) showed that this discrepancy is removed by allowing for the crystal field produced by the environment of the Yb^{3+} ions, which reduces their effective moment from $4\mu_B$ to $1{\cdot}7\mu_B$ in agreement with paramagnetic resonance experiments

† *Note added in proof* : The relative signs in this expression are correct only for ions in the second half of the rare earth series, but these comprise most of the cases found in practice. In the first half of the series the spin–orbit coupling is such that the diagonal components of $\mathbf{J}$ and $\mathbf{S}$ have opposite signs (Néel 1938), and the antiparallel alignment of the rare earth and iron *spins* then leads to a parallel alignment of their magnetic moments (Wolf 1961). Examples of this case are the mixed Nd–Y iron garnets, which have magnetic moments which are larger than that of pure YIG (Goldring, Schieber and Vager 1960).

(Boakes *et al.* 1959, White and Carson 1960). As a result of the crystal field the moment should be anisotropic, as in the case of $CoFe_2O_4$, and for certain rare earth ions large anisotropies are predicted. The effects of crystal fields on the rare earth ions in the garnets have also been considered by Ayant and Thomas (1959 b), Pauthenet (1959 b) and White and Andelin (1959). The special case of europium iron garnet has been considered by Wolf and Van Vleck (1960).

Another effect arises when there is a possibility of a re-arrangement of the ionic valencies among the different cations. This mechanism has been invoked by Harrison, Osmond and Teale (1957) and Eschenfelder (1958) to explain the anomalous moment of $MnFe_2O_4$. If the Mn ions are divalent and the Fe ions trivalent, as might be expected, the Néel theory predicts a moment

$$[5\mu_B + 5\mu_B] - (5\mu_B) = 5{\cdot}0\mu_B$$

compared with the observed value $4{\cdot}56\mu_B$ (Guillaud and Creveaux 1950). As both Mn^{2+} and Fe^{3+} ions are in S-states we cannot appeal to orbital effects to account for the difference in this case, but if there were some simultaneous formation of Mn^{3+} and Fe^{2+} a reduction of the moment could be explained. Evidence for this mechanism has recently been found by Miller (1959, 1960) and by Krupička and Závěta (1959). An alternative reason for the discrepancy has been suggested by Wickham (1959) who, following up a general discussion of Goodenough and Loeb (1955), suggested that the reduction was due to covalent binding of the Mn^{2+} ions in the rather small A sites (Mn ferrite is 80% normal). However neither of these two explanations seems to be entirely satisfactory in view of the ferrimagnetic resonance results of Dillon, Geschwind and Jaccarino (1955) and Tannenwald (1955 b) which indicate a g-value very close to 2, small anisotropy, and narrow linewidth. The magnetic moment of manganese ferrite is therefore still somewhat puzzling, but the two mechanisms discussed above will certainly apply in some cases.

There has recently been considerable discussion about the magnetic moments of various compounds containing chromium, many of which are anomalously low (Schindler *et al.* 1952, Gorter 1954, Lotgering 1956 b, Edwards 1959, Baltzer and Wojtowicz 1959, Jacobs 1959 b, see also Jacobs 1960 for further references). One possible explanation has been proposed by Baltzer and Wojtowicz (1959), who suggest that as a result of the large local distortions known to occur in some of these compounds (e.g. $NiCr_2O_4$, $CuCr_2O_4$) the effective spin of the Cr^{3+} ions is reduced from $\frac{3}{2}$ to $\frac{1}{2}$. The mechanism which they term 'spin quenching' appears to be similar to the reduction of the paramagnetic moment of, for example, Fe in $K_3Fe(CN)_6$, in which the interaction of the magnetic electrons with their surroundings is so strong that it overcomes the intra-ionic exchange coupling which normally makes the spin a maximum, subject to the Pauli principle. Van Vleck (1935) has shown that such a breakdown of Hund's rule can be due to either covalency or very strong crystal fields.

However even without going to such an extreme case large distortions from cubic symmetry can reduce the effective moment of an ion such as Cr^{3+}, as can be shown by the following simple example. For a spin $S = \frac{3}{2}$ the effect of an axial component superposed on a cubic crystal field can be represented by a term in the spin Hamiltonian of the form $D(S_z^2 - \frac{5}{4})$ (see for example Bleaney and Stevens 1953). If the cubic and axial fields were comparable without necessarily being

unusually large, D could be of the order of tens of cm^{-1}, or even larger, and the ground state would then be a spin doublet $\mathbf{S}' = \frac{1}{2}$. This would have an anisotropic g-tensor, with principal values of either $g_\parallel = 6, g_\perp = 0$ or $g_\parallel = 2, g_\perp = 4$, depending on the sign of D. In either case the mean saturation moment value would then be roughly of the order of $3\mu_B/2$, instead of the free spin value of $3\mu_B$, and it would be anisotropic.

A completely different mechanism for explaining anomalous moments follows from the observation of Yafet and Kittel (1952) that for certain ratios of the sublattice interactions 'triangular' spin arrangements as shown in Fig. 1(*e*) will have a lower energy than the antiparallel Néel arrangements. Their theory has been extended by Lotgering (1956 b), who applied it particularly to the problem of the chromites and concluded that this type of arrangement was consistent with the magnetic measurements on Mn, Fe and $NiCr_2O_4$. A similar conclusion was reached by Gorter (1954) for mixed spinels in the series $MnFe_{2-t}Cr_tO_4$ and by Edwards (1959) for $MnCr_2O_4$ and $MnCrAlO_4$. However neutron diffraction experiments on some of these compounds by Pickart and Nathans (1959) showed no sign of long-range ordering of the kind which would be implied by the simple form of the Yafet–Kittel–Lotgering theory, though short-range triangular order was not ruled out. On the other hand, neutron diffraction consistent with long-range triangular ordering has been reported for $CuCr_2O_4$ (Prince 1957 a) and also for Mn_3O_4 (Kasper 1959), both of which also show anomalously low moments. The position is therefore far from clear. Kaplan (1959, 1960 a) has recently pointed out that triangular arrangements are possibly not stable when the cation sub-lattices have very high symmetry (e.g. cubic), and distortions of the crystal structure may then be necessary to stabilize them. Empirically there seems to be no doubt that triangular arrangements can occur, but it appears that other effects may be more important in some cases. A particular system for which triangular ordering seems to be unambiguously predicted (de Gennes 1959) is $Y_{3-x}Ca_x[Sn_xFe_{2-x}](Fe_3)O_4$, for which Geller *et al.* (1960) have recently reported anomalous magnetic moments. So far no neutron diffraction experiments have been reported for this system. Evidence for triangular arrangements can also be inferred from high field susceptibility measurements (see § 3.3).

Another possible explanation of anomalous magnetic moments is the formation of a multi-spin-axis arrangement due to local anisotropy forces, as shown in Fig. 1(*f*). This would be most important when crystal distortions are large and exchange interactions are small. So far there has been no direct evidence for this type of order, which should be detectable by neutron diffraction.†

Yet another kind of mechanism which can lead to a change in magnetic moment is the possibility of variable cation population of the different types of lattice sites (Fig. 1(*d*)). This is the simplest of all the effects and the oldest (Néel 1948 a), but it is restricted to a few substances as most ions have very definite site preferences. It is generally very undesirable as it makes the magnetic moment dependent on the previous heat treatment, but it leads to the possibility of some elegant experiments in which measurements can be made on a single sample under differing conditions of site occupation and magnetization (see for example Rado and Folen 1957, also § 3.4).

† *Note added in proof*: Recent neutron diffraction experiments on $Ho_3Fe_5O_{12}$ by Herpin, Koehler and Mériel (1960) are briefly discussed in § 8.1.

Another way in which magnetic moments can be changed deliberately is by substitution of non-magnetic ions for some of the magnetic ones. A large number of such experiments have been reported, and it was a great success of the theory that it predicted an *increase* in the magnetic moment if substitution is made preferentially into the sub-lattices with the smaller magnetization (the A sites for ferrites). For small additions it is possible to calculate the change in the magnetic moment in terms of the moments of the displaced ions and good quantitative agreement is found (Gorter 1950, see also Gilleo 1960). We shall not repeat this classic discussion, which has been frequently reproduced, as it serves mainly to confirm the existence of ferrimagnetism about which there is now no doubt. More recently substitution experiments with diamagnetic ions have been made in the garnet and magnetoplumbite structures (Gilleo and Geller 1958, Villers, Pauthenet and Loriers 1959, Bertaut, Deschamps and Pauthenet 1959), with a view to relating cation preferences to structural considerations, and also as an aid to inferring sub-lattice structures (see for example Jacobs 1959 c). Substitutions are also frequently made for technical applications.

In compounds with more than two inequivalent sites the Néel theory is often used to infer possible sub-lattice arrangements from observed moments (Gorter 1954, 1957, see also Smit and Wijn 1959). In view of the many complications which have been observed in simpler compounds such analyses should only be considered as a first approximation, to be confirmed by substitution or neutron diffraction experiments, particularly when 'suspect' ions such as Co^{2+} or Cr^{3+} are present.

3.2.2. *Temperature dependence of magnetization.*

Néel's theory was also successful in predicting the variety of magnetization–temperature curves shown diagrammatically in Fig. 2. Some of these are illustrated by the results on rare earth iron garnets shown in Fig. 5 (Pauthenet 1958). So far no curve corresponding exactly to Fig. 2(*c*) has been observed, though the results for YbIG are very similar to it. From a theoretical point of view such a situation would be interesting as it could imply two similar sub-lattices cancelling identically at $T = 0°\text{K}$, i.e. pure antiferromagnetism, changing to ferrimagnetism at higher temperatures due to differences in interactions, but not intrinsic moments. In YbIG the cancellation is purely accidental.

The most striking feature of the various curves is the possibility of a compensation point as already mentioned in § 1. In order to observe the magnetization actually passing through zero at the compensation point it is necessary to use a very homogeneous sample, and the sharpness of the magnetization minimum may be used as a measure of the uniformity of the composition. The existence of compensation points was first demonstrated by Gorter and Schulkes (1953) in mixed $Li_{0\cdot5}Fe_{(2\cdot5-a)}Cr_aO_4$, with $1\cdot0 < a < 1\cdot6$, by showing that the remanent moment reverses direction as a function of temperature near 38°c. Correspondingly, the spontaneous magnetization as measured in a large field was found to have a sharp minimum at the same point, though its magnitude was of course always positive. Compensation points have also been found in a number of other mixed compounds, but so far the only pure substances in which they have been observed are the garnets and Co_2VO_4 (Menyuk, Dwight and Wickham 1960). Magnetic resonance phenomena in the vicinity of compensation points are discussed in § 5.5.

If the basic assumption of antiparallel sub-lattices is relaxed many more types of anomalous magnetization curves can occur. Lotgering (1956 b), extending the work of Yafet and Kittel, considered the case of a normal spinel with an arbitrary number of sub-lattices on the A and B sites, the interactions between which could be specified by five molecular field coefficients: $n(A_iB_j)$, $n(A_iA_i)$, $n(A_iA_j)$, $n(B_iB_i)$, and $n(B_iB_j)$. With this simplifying assumption he found that only four types of completely ordered arrangements can occur, as shown in Fig. 6(*a*) ((i) to (iv)) and that it is also possible for the moments on one type of lattice site to become

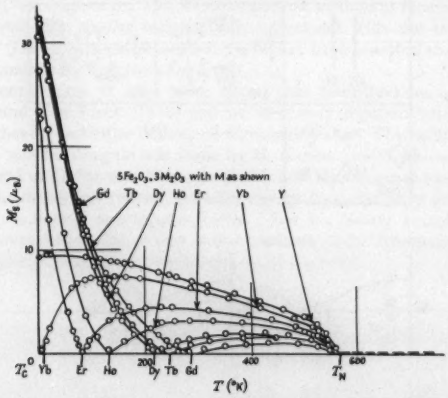


Fig. 5. Spontaneous magnetization–temperature curves for several rare earth iron garnets, in Bohr magnetons per gramme molecule ($5Fe_2O_3.3M_2O_3$) (*after* Pauthenet 1958).

ordered antiferromagnetically while the moments of the other type remain paramagnetic. These partially ordered states may be denoted by the symbols ↕∗ and ∗↕. Which of the six states occurs at a particular temperature will depend on the five interaction constants, and transitions between certain pairs of states are possible. The corresponding curves of magnetization against temperature are shown in Fig. 6(*b*). Another possible arrangement, forbidden by the Yafet–Kittel–Lotgering theory but suggested by the further calculations of Kaplan (1960 a), is one in which both the A and B sub-lattices split up, as shown in Fig. 6(*a*) (vii). There is, as yet, very little experimental evidence for any of these predictions but it is certain that multi-sub-lattice arrangements more complicated than the collinear Néel type alignment must occur in some cases. One important factor which has been neglected in the Lotgering analysis is the effect of anisotropy, which, as we have seen before, could be particularly important in some of the cases in which we might expect multi-axis alignment. The assumption of only five types of interaction may also not be adequate, but it is a rather necessary one in order to make the analysis tractable. In this connection some direct experimental information on the strength

of the exchange interaction between different pairs of moments might be most useful. The most stimulating experiments, however, would be unambiguous magnetic or neutron evidence for the different arrangements and the transitions between them.

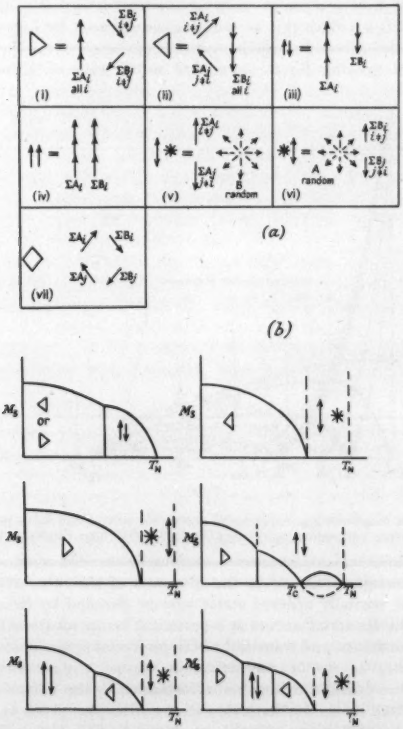


Fig. 6. (*a*) (i) to (vi) Illustration of the four ordered and two partially ordered sub-lattice arrangements predicted by the Yafet–Kittel–Lotgering theory (*after* Lotgering 1956 b); (vii) a further possible arrangement indicated by Kaplan's (1960 a) extension. (*b*) Spontaneous magnetization–temperature curves corresponding to some of the arrangements shown in (*a*), indicating the possible transitions between the various states of order (*after* Lotgering 1956 b).

For a simple two sub-lattice ferrimagnet the observed magnetization curve may be fitted to the Néel theory with the same three molecular field constants n, α and β, as used for the susceptibility above the Néel point. A convenient analytical procedure for finding the constants has been outlined by Rado and Folen (1960).

In general a very reasonable fit can be obtained (see for example Pauthenet 1952), but in spite of this the constants deduced should not be considered to be an accurate measure of the exchange interactions (see for example Smit and Wijn 1959). The main reason for this is that the curves, which are automatically made to fit at $T = T_N$ and $T = 0$, are not at all sensitive to the exact values of the parameters, and this conceals the relative inadequacy of the molecular field theory, particularly near T_N and absolute zero. This point was already recognized by Néel, but unfortunately even the simplest known ferrimagnetic structures are so complex that any theory statistically more exact than the molecular field approximation is very difficult. Valenta (1959) has adapted the Van Vleck expansion method to ferrimagnetism and obtained approximate results essentially in agreement with the molecular field calculations. The only improvement on this which has been tried so far is the spin wave approximation for low temperatures.

An excellent review of spin wave theory has recently been given by van Kranendonk and Van Vleck (1958) and we shall only mention here some recent papers which have a particular bearing on ferrimagnetism. The earliest application of spin waves to a ferrimagnet was made by H. Kaplan (1952) who used the semi-classical method of Heller and Kramers (1934) (see also Herring and Kittel 1951) to calculate the dispersion relation between energy $\hbar\omega_k$, and wave vector $\mathbf{k}$ for the spin waves in a normal spinel-type ferrite. For the lowest branch he found a quadratic relation $\hbar\omega_k = Dk^2$, where D is a constant, as for ferromagnetism. This leads to a magnetization–temperature relation of the form

$$M_T = M_0(1 - aT^{3/2} + \ldots) \qquad \ldots\ldots(3.3)$$

where a depends on the three exchange constants J_{AB}, J_{AA}, J_{BB} (cf. J_{ij} in Eqn (2.1)) and the number of nearest neighbours. For the normal spinel structure the constant a is given by

$$a = \frac{0{\cdot}0587}{|\mathbf{S}_A + 2\mathbf{S}_B|}\left[\frac{k_B T|\mathbf{S}_A + 2\mathbf{S}_B|}{|11J_{AB}\mathbf{S}_A\mathbf{S}_B/2 + J_{AA}S_A^2 + 2J_{BB}S_B^2|}\right]^{3/2} \qquad \ldots\ldots(3.4)$$

where $\mathbf{S}_A$ and $\mathbf{S}_B$ are the spins on the A and B sites, and k_B here stands for the Boltzmann constant.

At high energies there is a second branch in the dispersion relation which has the form $\hbar\omega_k = D' + D''k^2$, but this is unimportant for the low temperature magnetization, since $D' \sim J \gg k_B T$. Qualitatively similar results were later found independently by Kondorskii, Pakhomov and Skiklosh (1956).

In striking contrast to these results Vonsovski and Seidov (1954), using a strictly quantum mechanical approach, found a linear dispersion law $\hbar\omega_k = Ck$, for the lowest branch, and this disagreement was not resolved until 1958, when T. Kaplan re-examined the problem. He found that the difference between the two results was not due to any basic distinction between quantum and semi-classical mechanics, which in this problem lead effectively to the same result, but rather to the method of averaging over the two sub-lattices. He pointed out that both groups of authors had made the assumption that the spin-wave amplitudes were the same for all the spins on any one sub-lattice, but whereas Kaplan had used this condition in the equation of motion, Vonsovski and Seidov had taken the average already in the Hamiltonian. To avoid this difficulty he repeated the

calculation without making this assumption at all, and he used the rigorous quantum mechanical method of Holstein and Primakoff (1940). This calculation showed that H. Kaplan's method was essentially correct, both in its method of averaging and also in the treatment of the spin as a semi-classical vector when S is large. Unfortunately in order to simplify the calculation he neglected J_{AA} and J_{BB} compared with J_{AB}, so that the final result, which is otherwise exact for a finite spin, does not correspond to any real ferrimagnet. Even so it now seems clear that ferrimagnetic spin waves are very similar to ferromagnetic ones, as might be expected naïvely, with ω proportional to k^2 and $M_0 - M_T$ proportional to $T^{3/2}$. So far the coefficient a has only been calculated in the classical limit and only for the normal spinel structure,† but in the absence of data on the J's it must in any case be treated purely as an empirical parameter. Refinements to the theory have recently been made for the region of long wavelength spin waves (small k) and the effects of applying a magnetic field. These will be discussed in connection with resonance line widths in § 5.7. Spin waves are also important for low temperature specific heats and thermal conductivities (see §§ 4.1 and 4.2) and neutron scattering experiments (see § 6.2).

In the past there have been very few attempts to compare magnetization data with Eqn (3.3), mainly because it was thought that the spin wave treatment was valid only up to temperatures which are a small fraction of T_N. The few measurements which had been made at very low temperatures were not really accurate enough (on account of troubles from anisotropy, impurities, etc.) to distinguish between a $T^{3/2}$ law and some other gradual approach to M_s. Pauthenet (1952) concluded that $MnFe_2O_4$ obeys Eqn (3.3) up to $T/T_N \sim 0{\cdot}7$, but that in the cases of Ni, Co and Fe^{2+} ferrites a better fit was obtained with a law of the form $M_0(1-bT^2)$. It may be significant both from an experimental and theoretical point of view that $MnFe_2O_4$ has a much smaller anisotropy since it contains only ions in S-states (see § 3.4).

3.3. *Susceptibilities below the Néel Point*

In the previous section we discussed spontaneous magnetization as an intrinsic property without any reference to how it was determined in practice. As in the case of ferromagnets, little or no spontaneous magnetization can generally be observed in the absence of an applied field, due to the formation of domains. The removal of these by a field gives rise to the complex phenomena of ' technical saturation '. These have been discussed extensively for ferromagnetism (see for example Bozorth 1951) and are of no special interest from the point of view of the theory of ferrimagnetism, but the effects are important in the practical determination of the true saturation magnetization. For polycrystalline samples it is generally found that the magnetization rises at first rather rapidly with applied field and then approaches a limiting value more slowly, varying finally as $1/H$ or $1/H^2$ (see for example Néel *et al.* 1960). The experimental procedure for finding the saturation then consists in plotting the measured M against $1/H$ and extrapolating the resulting line or curve to zero. This does not imply that the value so obtained really corresponds to M in very large fields, since it is assumed throughout that

† *Note added in proof* : Meyer and Harris (1960) and Douglass (1960) have recently reported a similar calculation for YIG, and Dreyfus (1960) an extension to the case of gadolinium iron garnet.

the applied fields are small compared with the exchange fields which maintain the ferrimagnetic order. For sufficiently large fields one would always obtain the ferromagnetic moment corresponding to all spins parallel, but generally this would require the application of several million gauss. The mechanisms which control the form of the field dependence are very complex (Holstein and Primakoff 1941, Néel 1948 b), but generally a large crystalline anisotropy leads to a slow approach to saturation. These effects are eliminated almost entirely by using a single crystal with its ' easy ' axis parallel to the measuring field.

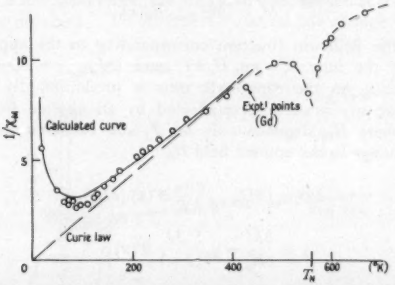


Fig. 7. Typical susceptibility–temperature curve below the Néel point for a substance containing a ' loosely coupled ' sub-lattice. Experimental points from Pauthenet (1958). The full line curve is calculated from Eqn (3.6) and Pauthenet's values of M_{Gd} (W. P. Wolf, unpublished). The anomaly at high temperatures is due to a contribution from the Fe^{3+} sub-lattices approaching their Néel point (Pauthenet 1958).

In addition to these effects the magnetization may in certain cases show an intrinsic dependence on field, and this can be described by a field independent susceptibility χ. This susceptibility may be due to at least one of four effects. The most trivial of these arises near the Néel point from the fact that the sub-lattices are only partially ordered, so that a relatively small field can produce a change in the moment, as above the Néel point. Even at temperatures well below the Néel point this can give rise to a contribution to χ which may not be entirely negligible. So far this effect does not seem to have been studied in detail, mainly because in polycrystals, on which most magnetization work has been done, technical saturation effects complicate the interpretation. Evidence for this type of susceptibility has been obtained in the rare earth iron garnets (Pauthenet 1958) as can be seen in Fig. 7, which shows the susceptibility for the Gd compound.

In addition to the anomalous behaviour in the region of the Néel point these compounds show another effect, which in fact dominates the susceptibility over a wide range of temperature. This arises from the fact that the rare earth ions are coupled only loosely to the iron sub-lattices, so that the susceptibility mechanism discussed above will be important for these ions even at temperatures far below the Néel point of the whole structure. To a first approximation, the effective molecular field, H_e, acting on the rare earth ions is so small (Pauthenet 1958) that χ is approximately equal to the ionic susceptibility in the absence of any exchange

interactions. If crystalline field effects are also small, as they certainly are for Gd^{3+}, χ should then follow Curie's law, in close agreement with observed values for GdIG shown in Fig. 7. The small deviations from Curie's law down to 100°K were explained by Pauthenet in terms of an antiferromagnetic interaction between the rare earth ions, but this has subsequently been shown to be small (Cohen and Ducloz 1959, Pauthenet 1959 a, b), and a much more important correction arises from paramagnetic saturation due to the finite value of H_e (Wolf and Van Vleck 1960). In terms of the molecular field theory, the sub-lattice magnetization M_s is given by

$$M_s = (M_s)_{T=0} B_J(x), \qquad \ldots\ldots(3.5)$$

where $B_J(x)$ is the Brillouin function corresponding to the appropriate angular momentum J of the ions, $x = g\mu_B H_e/kT$, and $(M_s)_{T=0} = Ng\mu_B J$. Since the effective field acting on the rare earth ions is predominantly due to the Fe sub-lattices, which are practically unaffected by an applied field, we can take $H_e = H_{Fe} + H_0$, where H_{Fe} depends only on T, and calculate the change in M produced by a change in the applied field H_0

$$\delta M = (M_s)_{T=0} \frac{g\mu_B}{kT} B'(x)\,(\delta H_0), \qquad \ldots\ldots(3.6)$$

so that

$$\chi = \frac{\delta M}{\delta H_0} = \chi_C \left[\frac{3J}{J+1} B'(x)\right],$$

where χ_C is the Curie law susceptibility. When $kT \gg g\mu_B H_e$, $B'(x) \to (J+1)/3J$, and $\chi \to \chi_C$ as expected, but at low temperatures $\chi \to 0$ as $B'(x) \to 0$. At any finite temperature the value of $B'(x)$ may be estimated directly from the sub-lattice magnetization by eliminating x between Eqns (3.5) and (3.6). Using Pauthenet's deduced values of M_s for the Gd sub-lattice in GdIG, this leads to the calculated curve shown in Fig. 7, without any further assumptions. Qualitatively the same considerations will also apply to the other rare earth garnets, but for these the effects of crystalline electric fields may also be important (Wolf and Van Vleck 1960). Susceptibility effects of this type will be large whenever one or more sub-lattices are loosely coupled to the others, so that they are far from saturation even at low temperatures. A similar effect can be produced by paramagnetic impurities in normal ferrimagnetic substances, if the impurity is present as a second phase which is weakly coupled to the main moment.

Another mechanism which can give rise to a susceptibility, even at very low temperatures, is due to the same effect as that which produces temperature independent paramagnetism in weakly interacting systems (Van Vleck 1932). If any of the ions present have a ground state which in the combined exchange and crystal fields has matrix elements of the magnetic moment with any of the higher states, these will give rise to a low temperature susceptibility

$$\chi = N \sum_i \frac{2b_{oi}^2}{E_i},$$

where b_{oi} are the matrix elements and E_i the energies of the excited states. In general this effect will be small, but it may become appreciable if some of the E_i are small (e.g. $\chi_M \sim 10^{-3}$ for $E_i \sim 100k$) and also if measurements are made in very

high fields. So far this effect has not been considered in detail and the only high field magnetic moment measurements which have been reported (Jacobs 1959 a, b, c, 1960) have been interpreted in terms of another mechanism, which may be more important in certain cases.

Jacobs pointed out that the magnetic moment of a ferrimagnet with a multi-axis type of ordering is sensitive to an applied field, since the angle between the sub-lattices can be changed even at absolute zero. The change which can be induced depends on the relative strengths of the applied and the exchange fields, and for substances with a low Néel point (say 50°K), an effective molar susceptibility of the order of 10^{-2} is obtained. To observe an appreciable change in moment, large

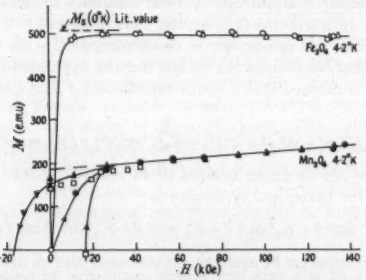


Fig. 8. Magnetization measurements in very high fields showing the contrast between a substance readily saturated (Fe_3O_4) and one which has an appreciable high field susceptibility (Mn_3O_4) (*after* Jacobs 1959 c).

fields, of the order of 100 k gauss, are required (Fig. 8). Applying these considerations to the compounds $MO.Mn_2O_3$, with M = Mn, Co and mixtures of these with Zn and Mg, Jacobs was able to relate the low saturation moments which had been observed with the susceptibilities measured by a pulsed field method (Jacobs and Lawrence 1958). He concluded that these compounds order in a triangular Yafet–Kittel arrangement, with Mn^{3+} ions on the B sites in 2 (or 4) non-parallel sub-lattices, their resultant being antiparallel to the M^{2+} ions in the A sites. In the case of $M = Mn^{2+}$ this conclusion has been supported by subsequent neutron diffraction experiments (Kasper 1959). (See however § 8.1.)

Similar high field susceptibility measurements could also be made on compounds in which multi-axis ordering arises from local anisotropy and in some cases it may not be easy to distinguish between the two. In general the possibility of a contribution from off-diagonal matrix elements must also be considered.

3.4. *Crystalline Anisotropy*

For ferromagnetic substances the crystalline anisotropy is defined as that part of the total free energy F which depends on the orientation of the magnetization vector **M** relative to the crystal, and this definition may be carried over to ferrimagnetism if **M** is taken to mean the resultant of the various sub-lattices. However

in this form the definition conceals the basic origin of the anisotropy, which in the ferrimagnetic case depends on the magnitudes and directions of the individual sub-lattice magnetic moments, rather than on their resultant. Thus if we express the total free energy formally as a sum over sub-lattices

$$F = \sum_{i,j} F_i(\mathbf{M}_i, \mathbf{M}_j),$$

each of the sub-lattice anisotropy functions F_i depends on the local symmetry at the lattice site of the ith type, whereas the overall crystalline anisotropy energy F is governed by the symmetry of the crystal structure as a whole. By considering only this total anisotropy it is therefore possible to obtain a phenomenological description of the measured anisotropy without reference to the detailed structure, and this is the procedure which has generally been followed. This does not lead to any difficulties when the anisotropy is small compared with the energy of the ferrimagnetic alignment, as each of the F_i can then be expanded in terms of simple series, which can be combined into a series expansion for F. For cubic symmetry this has the form

$$F = \text{const} + K_1(l^2 m^2 + m^2 n^2 + n^2 l^2) + K_2(l^2 m^2 n^2) + \mathrm{O}(l, m, n)^8, \quad \ldots\ldots(3.7)$$

where l, m and n are the direction cosines of $\mathbf{M}$ relative to the cubic axes. The corresponding form for hexagonal symmetry is

$$F = \text{const} + K_1' \sin^2\theta + K_2' \sin^4\theta + K_3' \sin^6\theta + K_4' \sin^6\theta \cos 6\phi, \quad \ldots\ldots(3.8)$$

where θ is the angle between $\mathbf{M}$ and the crystal axis, and ϕ defines the direction in the $\theta = \frac{1}{2}\pi$ plane.

In many cases the leading terms in these expansions can be fitted very closely to experimental results, which can then be expressed very simply in terms of one or two K's. However in some cases, and particularly at low temperatures, higher order terms may become comparable with the first, and convergence of the series is then poor. One reason for this may be that the local anisotropy of one of the sub-lattices is big or comparable with the exchange energy, so that a series expansion of F_i may be a very poor approximation for its angular variation. It is of course always possible to expand any angular function in terms of an infinite series of polynomials of the direction cosines having the appropriate overall symmetry, but it may sometimes be much more convenient to derive a completely different form. For example Slonczewski (1958) showed that small amounts of Co^{2+} in Fe_3O_4 will give rise to cubic anisotropy of the form

$$F = A \sum_i \log\cosh(B\cos\theta_i), \quad \ldots\ldots(3.9)$$

where θ_i is the angle between $\mathbf{M}$ and one of the four [111] axes and A and B are constants. Similarly Wolf (1959) showed that certain rare earth ions in the garnet structure may at low temperatures give rise to a cubic anisotropy of the form

$$F = \sum_{x=l,m,n} |[A^2 + (B^2 - A^2)x^2]^{1/2}|, \quad \ldots\ldots(3.10)$$

which in extreme cases could reduce to

$$F = B(|l| + |m| + |n|). \quad \ldots\ldots(3.11)$$

These expressions illustrate the fact that the normal expansions (3.7, 3.8) are by no means the only possible forms, though in practice they are often the most convenient. In general, if data can be fitted with $K_2 \ll K_1$, other higher order terms will probably be negligible ; if the initial convergence is poor an alternative form is indicated.

3.4.1. *Experimental methods.*

Anisotropy in the free energy can be deduced from three main kinds of experiments : (i) from the torque on a crystal in a field strong enough to produce technical saturation (see for example Bozorth 1951, Penoyer 1956), (ii) from the approach to saturation of the magnetization for various orientations of a single crystal (see for example Smit and Wijn 1959, Néel *et al.* 1960) and (iii) from the variation of the resonance field in a microwave experiment. Of these (i) and (ii) measure directly the static anisotropy which is the thermodynamic function as defined above. Microwave experiments measure the true anisotropy only under certain conditions, which are discussed further in § 5.4. In all methods it is usual to measure an angular variation in some principal plane and to analyse this into components of the form of Eqn (3.7) or (3.8). For this purpose the variation in certain planes is very much more sensitive to particular K's (see for example Sato and Chandrasekhar 1957), and a cross check between several planes should be made. In certain cases complex and irreproducible variations may be observed and these require special interpretation (Smit, Lotgering and Enz 1960).

3.4.2. *Theory.*

The combination of sub-lattice effects can give rise to a considerable variety of anisotropy–temperature curves, in a way similar to Néel's theory for the spontaneous magnetization. Some of the types of curve possible for a two sub-lattice case are illustrated diagrammatically in Fig. 9. It is clear that the temperature variation of the anisotropy in general bears no simple relation to the total magnetic moment as it does in the ferromagnetic case, though the contribution from each sub-lattice may depend simply on its own magnetization. Effects due to different sub-lattices were first observed by Folen and Rado (1956, 1958) (see also Rado and Folen 1956), who carried out anisotropy measurements on magnesium ferrites which had been subjected to different forms of heat treatment. As discussed in §§ 2.2 and 3.2, this causes a variation in the relative numbers of magnetic ions on the A and B type sites, and so also of the different contributions to the total anisotropy. For each sample the anisotropy was found to be small and of the first order only (K_1), and at any temperature it was possible to express K_1 in terms of two sub-lattice constants K_{1A} and K_{1B}

$$K_1 = n_A K_{1A} + n_B K_{1B} \qquad \ldots\ldots(3.12)$$

where n_A and n_B are the numbers of ions on the two sites, as deduced from the magnetization. Similar experiments have been carried out more recently by Belson and Kriessman (1959 a), and also by Pearson (1959 a), who used substitutions of diamagnetic Al and Ga in Fe_3O_4 to control the relative numbers of magnetic ions on the two sub-lattices.

The interpretation of the above experiments contains the implicit assumption that the individual sub-lattice anisotropies are really independent of one another,

i.e. that F_i depends only on $\mathbf{M}_i$. This will be true whenever the anisotropy arises from an interaction of the magnetic ions with their immediate (non-magnetic) surroundings, rather than from an anisotropic coupling between magnetic neighbours, and this seems to be the case in many ferrimagnetic crystals (Yosida and Tachiki 1957, Wolf 1957). However, in general the origin of anisotropy is a very complex question, and each type of compound must be considered separately. The big simplification in this problem, as compared with the theory of the ferromagnetic metals, is that it is reasonable to think in terms of localized magnetic ions behaving much as in paramagnetic salts, for which there is a good theory and much experimental data from microwave resonance.

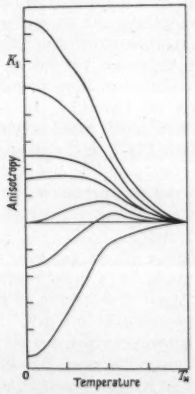


Fig. 9. Illustration of various curves of anisotropy (first-order cubic, K_1) against temperature possible for a simple two sub-lattice ferrimagnet (*after* Wolf 1957).

Although the exact mechanism of the indirect exchange interaction has not yet been determined, it seems to be generally agreed that it can be expressed to a good approximation in the form of Eqn (2.1), which is completely isotropic with respect to the orientation of the magnetic moment. With this assumption, anisotropy must arise from other interactions, the main ones of which can be expressed for each ion, formally, as

$$V_{\text{ion}}+V_{\text{cryst}}+V_{\text{so}}+V_{\text{ss}}+V_{\text{dd}}, \qquad \ldots\ldots(3.13)$$

where V_{ion} determines the electronic state of the free ion, V_{cryst} the influence of environment on the orbital motion, V_{so} the spin–orbit coupling $\lambda\mathbf{L}.\mathbf{S}$, V_{ss} the *intra*-ionic spin–spin coupling (see Pryce 1950, Abragam and Pryce 1951), and V_{dd} the *inter*-ionic dipole–dipole coupling. According to the relative importance of these terms we can distinguish several different cases.

(i) $V_{\text{ion}} \rightarrow$ S-state ion : e.g. (Mn^{2+}, Fe^{3+}, Gd^{3+}).

In this case the anisotropy will be small, as the combined effect of the three terms V_{cryst}, V_{so} and V_{ss} in Eqn (3.13) is non-zero only in a very high order of perturbation, with free ion states at very high energies (see for example Watanabe 1957). Without enquiring into the detailed mechanisms, one can construct a spin Hamiltonian from a knowledge of the crystal field symmetry (Bleaney and Stevens 1953), and express the anisotropic interactions in terms of a set of empirical constants, which could in principle be determined from paramagnetic resonance experiments on isolated ions in a similar environment. For $S = 2$ or $\frac{5}{2}$ the cubic term is $a(S_x^4 + S_y^4 + S_z^4)$ to which an axial distortion adds terms of the form $DS_\zeta^2 + FS_\zeta^4$ where ζ denotes the direction of the axis. For $S = \frac{7}{2}$, corresponding to Gd^{3+} (and Cm^{3+}) there are additional terms $a'(S_x^6 + S_y^6 + S_z^6) + qS_\zeta^6$. In the general case there may also be various rhombic terms, and terms in $S_\pm^4$ and $S_\pm^6$. In paramagnetic compounds it is usually found that $D \gg a \sim F$. The effect of anisotropic terms of this type on the ferrimagnetic anisotropy has been considered by Yosida and Tachiki (1957) and by Wolf (1957). In cubic compounds the contribution from the large term in D vanishes identically in first order and the largest anisotropy comes from a. For each type of term the temperature dependence of the contribution to the anisotropy energy may be calculated in the molecular field approximation from $\langle \mathbf{S} \rangle_{\text{av}}$, which is proportional to the sub-lattice magnetization $\mathbf{M}_i$. Using this theory and assuming that only the term in a is important Yosida and Tachiki were able to fit the measured anisotropy of $MnFe_2O_4$ (Tannenwald 1955 b) over a wide temperature range, by choosing two values of a for the Fe^{3+} ions on A and B sites and $a = 0$ for Mn^{2+}.

The magnitudes of the parameters a were found to be very reasonable, but at the time one puzzling feature was that both values of a were found to be positive, whereas crystal field theory (Van Vleck and Penney 1934) had predicted opposite signs for tetrahedral and octahedral coordination. However, the more extended theory of Watanabe (1957) now indicates that a would in fact be positive for either type of coordination,† and this seems to be confirmed by paramagnetic resonance (Geschwind 1959). On the other hand, Folen and Rado (1958) applying the same theory to interpret the temperature variation of K_{1A} and K_{1B} for $MgFe_2O_4$ (Eqn (3.12)) deduced values of a_A and a_B which had reasonable magnitudes but had opposite signs. The difference between these two cases may lie in the fact that $MnFe_2O_4$ is an almost normal spinel in which the local symmetry is predominantly cubic, while in the partially inverse spinel $MgFe_2O_4$ local distortions may make the term in F or the second-order contribution from D more important. A similar conclusion has recently been reached by Folen (1960), who compared the anisotropies of a sample of $Li_{0.05}Fe_{2.45}O_4$ under different conditions of long-range cation order (varied by suitable heat treatment). The importance of the quartic term in F has also been indicated by the paramagnetic resonance experiments of Geschwind (1959, 1961), which showed that even in a well-ordered structure such as the garnets there may be large local deviations from cubic symmetry, which give rise to terms in F comparable with those in a. A suitable combination of these

† *Note added in proof* : Further calculations by Powell, Gabriel and Johnston (1960) have recently shown that a is not *necessarily* positive, though it is positive over a wide range of parameters. At present the theoretical position on this point is therefore not quite conclusive.

has been invoked by Rodrigue, Meyer and Jones (1960) to account for the observed anistropies of YIG and GdIG, and in the latter case they found evidence also for a second-order term in D.

(ii) $V_{\text{ion}}+V_{\text{cryst}} \rightarrow$ 'effective S-state' (e.g. Cr^{3+}, Ni^{2+} in octahedral symmetry, Co^{2+} in tetrahedral symmetry, and also some other ions in lower symmetry).

In several cases a strong crystal field can effectively remove the orbital degeneracy of an ion for which $\mathbf{L} \neq 0$, and lead to an 'effective S-state'. Such an ion can be described in much the same terms as above, except that the basic mechanisms which give rise to the spin Hamiltonian will be different. Most of the transition group ions of this type correspond to $S<2$ and the fourth-order terms in a and F therefore vanish identically. On the other hand D will in general be rather larger, since it now arises from perturbation with states only at crystal field energies, rather than with higher spectroscopic terms. In cubic compounds the anisotropy from this origin is still rather small since it arises only in a second approximation, but even then it can be appreciable if the exchange interaction is weak or local distortions are large. The effect will also be large when there are low-lying orbital states as a result of an otherwise degenerate ground state split by a relatively small distortion.

This case apparently arises for the Fe^{2+} ions in Fe_3O_4, for which an octahedral cubic field gives rise to an orbital doublet ground state which is split by the local trigonal symmetry. However no direct measurements of D, a or F are available, and Yosida and Tachiki, who considered the problem theoretically, concluded that there were several competing processes which make an estimate uncertain. They showed that the relatively small anisotropy observed experimentally is reasonable, but their calculations have been criticized by Van Vleck (1959) who pointed out that the problem was even more complicated than they had assumed.

In addition to the effects discussed above the proximity of excited orbital states may give rise to any one of four other mechanisms, all of which have sometimes been termed 'anisotropic exchange'. The first of these was considered in detail by Yosida and Tachiki for the case of Ni^{2+} in $NiFe_2O_4$. It arises from the fact that an isotropic exchange interaction acting equally on all spin states of a given ion, regardless of orbital state, can change the effect of the coupling of excited states to the ground state by the spin–orbit interaction. This is a complicated effect which only contributes to cubic anisotropy in a high order, and Yosida and Tachiki showed that the corresponding K_1 was proportional to $\lambda^4 J^2/E_0^5$, where E_0 is the energy of an excited orbital state, which for Ni^{2+} is of the order 10^4 cm^{-1}. A numerical estimate showed that in this case the resulting anisotropy is quite negligible, but it could be important if E_0 was smaller.

A similar and generally larger effect can arise in a lower order if the exchange interaction parameter J is not independent of the orbital state. Van Vleck (1951) showed that this can give rise to a cubic anisotropy of order $\lambda^4 J/E_0^4$ which may be expressed in terms of $(g-2)^4 J$, where g is the spectroscopic splitting factor, whose deviation from 2 also depends on the spin–orbit coupling. So far this mechanism has not been applied in detail to any ferrimagnetics.

In one sense neither of these two interaction effects is really an anisotropic coupling, since they both depend only on the orbital states of one ion, with the

neighbours merely providing an isotropic exchange field. Greater complications arise when both neighbouring ions have low lying orbital states, and this can give rise to 'pseudo-dipolar' and 'pseudo-quadrupolar' coupling, first proposed by Van Vleck (1937) to explain the anisotropy of ferromagnetic metals. Both these effects probably contribute appreciably to the anisotropy in certain cases (Stevens 1953), but so far no one has attempted to consider them in detail for any ferrimagnetic compound.

Closely related to the possible dependence of J on the orbital states of the interacting ions is the reasonable certainty that the basic indirect exchange mechanism is itself direction dependent and not quite of the form $J_{ij}\,\mathbf{S}_i.\mathbf{S}_j$. For example, Moriya (1960) considering the general superexchange mechanism in the presence of spin–orbit coupling found an additional term of the form $\mathbf{D}.[\mathbf{S}_i \times \mathbf{S}_j]$. Such an anisotropic term could in certain circumstances dominate all the other effects, and a clarification of the whole position must await further refinement of the theory and direct experimental evidence.

(iii) $V_{\text{ion}} + V_{\text{cryst}} \rightarrow$ degenerate orbital multiplet (e.g. Co^{2+} in octahedral and trigonal symmetry, Ni^{2+} in axially distorted tetragonal symmetry).

If the symmetry of the crystal field is sufficiently high it may in certain circumstances not remove all the orbital degeneracy, and this can produce an extremely large anisotropy, over 100 times bigger than in the previous cases. This is because the spatial anisotropy of the orbital states is then coupled directly to the spins by the spin–orbit interaction, and not as a result of a high order perturbation effect. Details of this type of anisotropy have been worked out by Slonczewski (1958), who was able to account with some success for the magnitude and temperature dependence of the anisotropy produced by small additions of Co^{2+} to Fe_3O_4, as measured by Bickford, Brownlow and Penoyer (1957). In this system local trigonal symmetry produces a strong axial anisotropy in the Co^{2+} ions, the axis lying along one of the four [111] directions, and a combination of these gives rise to the unusual form of the cubic anisotropy indicated in Eqn (3.9). However this mechanism is unlikely to be important in many other cases since it requires a relatively high local symmetry, which is easily destroyed by irregularities such as occur in inverse spinels. As the degeneracy is gradually removed there is a transition to the types of mechanisms discussed in previous sections and a smaller anisotropy results (see for example, Bickford, Brownlow and Penoyer 1957, Pearson 1959 a, b 1960, Kaminow 1960).

(iv) $V_{\text{cryst}} \sim V_{\text{ion}}$.

It has been suggested by Smit (1959) that the large axial anisotropy observed in $BaFe_{12}O_{19}$ (Went *et al.* 1952) arises, in part, from a small number of Fe^{3+} ions (1 in 12) which are bound so strongly by their environment that the free ion S-state is uncoupled to give an orbitally degenerate ground state like that of Fe^{2+}. This mechanism is similar to 'spin quenching' discussed above in connection with low magnetic moments, and it is suggested in this case by the fact that the five-fold coordinated sites which accommodate 1/12 of the Fe ions are small compared with their ionic radius. No details of the calculations have been given, but it is clear that while this effect can be important in a few tightly bound compounds it will not apply in general.

(v) V_{dd} large.

Another important contribution to the anisotropy of $BaFe_{12}O_{19}$ arises from classical magnetic dipole–dipole coupling (Smit and Duyvesteyn 1959). This effect will in general be large in all compounds with non-cubic overall symmetry, and particularly those with big atomic magnetic moments such as Fe^{3+}. To calculate the magnitude of the anisotropy it is only necessary to know the sub-lattice arrangement and the size of the atomic moments, whose mutual interaction energies are then summed over the crystal lattice by direct computation. It is generally assumed that the magnetic moments are free (i.e. isotropic), corresponding to conditions (i) or (ii) above, but anisotropic moments such as would be implied by (iii) or (iv) could also be taken into account. In cubic crystals this effect vanishes identically in first order by symmetry, and in second order it is small (Van Vleck 1937, Yosida and Tachiki 1957) unless the exchange interaction is unusually low, i.e. the Néel point is of the order of a few degrees Kelvin.

(vi) $V_{so} \gg V_{cryst}$.

This condition holds for all the magnetic rare earth ions except Gd^{3+}, and it implies that **L** and **S** are so strongly coupled that the crystal field acts only on their resultant **J**. Due to shielding by outer electrons the effect of the crystal field is small compared with that in the 3d group and roughly comparable with the exchange interaction. This presents a complicated problem which has not yet been solved, mainly because there are not enough experimental data to fix the necessary parameters. In general the contribution to the anisotropy will be very large and it will reflect strongly the local symmetry at the rare earth sites. This may give rise to unusual forms of the overall anisotropy as illustrated in Eqns (3.10) and (3.11). Kittel (1960) has recently pointed out that a particularly extreme form of anisotropy occurs if the two lowest energy levels in the combined crystal and exchange fields intersect at some orientation of **M**. At $T = 0°\text{K}$ this produces an abrupt change in the free energy at the critical angle, and even at a finite temperature the angular variation of F cannot be expanded in terms of any simple form. It is likely that some of the anomalous microwave resonance results on impure yttrium iron garnet at low temperatures arise from this effect (see § 5.4).

We may summarize this section with the general conclusion that the main features of the theory of ferrimagnetic anisotropy are now well understood in terms of the paramagnetism of localized magnetic ions, although many details have not yet been worked out. A complete treatment, however, will require very careful consideration of many individual cases.

3.4.3. *Macroscopic effects.*

The signs and magnitudes of the various anisotropy coefficients determine the directions of minimum and maximum energy, the so called ' easy ' and ' hard ' axes of magnetization, and an interesting situation has been observed in several of the complex hexagonal magnetic oxides (§ 2.2(*d*)). Smit (1959) has pointed out that if $-2K_2' < K_1' < 0$, and other K's are negligible, there is no single stable direction for the magnetization vector but rather an ' easy cone ' making a semi-vertical angle $\theta = \sin^{-1}(-K_1'/2K_2')^{1/2}$ with the axis. This effect leads to a reduced spontaneous moment when a field is applied along the axis and this has been

observed experimentally. Other hexagonal oxides have an 'easy plane' of magnetization, corresponding to the condition $K_1' < 0$; $|K_1'| \gg$ other K's. Within the plane the anisotropy $|K_4'|$ is very small and as $|K_1'|$ is generally very large these substances combine certain features of magnetically 'hard' and 'soft' materials. In particular they have large initial permeabilities up to high frequencies (see § 5.9.3) and they are therefore of considerable technical importance. A review of the general properties of these materials has been given by Smit and Wijn (1959).

These authors also give a good discussion of the concept of an 'effective anisotropy field' which is very often used to describe the anisotropy energy in the vicinity of a principal axis, though as a general rule it is safer to make deductions regarding torque, direction of magnetization, etc., directly from the free energy. They also present a summary of the experimental results of all anisotropy measurements which have been reported up to the present.

3.5. *Magnetic Annealing*

In the course of measurements of magnetic anisotropy it is sometimes observed that an otherwise cubic crystal can permanently acquire an additional axial anisotropy, as a result of being subjected to a magnetic field at an elevated temperature. The effect depends in a complex way on the magnitude and direction of the applied field, the temperature and the time of annealing, and it has been studied extensively in a number of ferrimagnetic compounds. There seem to be two general requirements for the effect to become observable : the substance must contain cobalt, and it must deviate from perfect stoichiometry, suggesting that Fe^{2+} or vacancies are also involved. So far four types of different mechanisms have been suggested to explain specific results.

(*a*) *Precipitates, possibly of* Fe_3O_4. Bozorth, Tilden and Williams (1955) suggested that elongated precipitates may be formed by migration of ions at high temperatures, and through magnetostriction these could influence the anisotropy. Experimental evidence for precipitates was found by Williams, Heidenreich and Nesbitt (1956) who used magnetic torque and electron diffraction measurements (see however Chikazumi 1956). Magnetization data on heat treated $CoFe_2O_4$ by Foner and Artman (1958) are also consistent with domains pinned by preferentially oriented inclusions.

(*b*) *Vacancies.* From a study of the annealing of cobalt ferrite as a function of Fe and oxygen concentration Iida, Sekizawa and Aiyama (1955) concluded that vacancies may become ordered in a short range manner, and then interact anisotropically with the Co^{2+} ions through spin–orbit coupling. A general review of the role of vacancies has been given by Iida (1960).

(*c*) *Pairs of ions.* A rather similar mechanism had previously been suggested by Néel (1954) and by Taniguchi and Yamamoto (1954) who pointed out that if pairs of Co^{2+} ions were formed by migration they could couple anisotropically to give a preferred direction. To make the migration possible at reasonable temperatures some vacancies would also be required.

(*d*) *Preferential population of locally anisotropic sites.* The simplest and most unambiguous studies of annealing have been made by Penoyer and Bickford (1957) who investigated pure Fe_3O_4 containing small amounts of cobalt. Their results were explained quantitatively by Slonczewski (1958), using the same theory as

that with which he had accounted for the cubic anisotropy. He had shown that the local symmetry produces a strong axial anisotropy along each of the four [111] directions, the number of Co^{2+} ions of each type being the same for cubic overall symmetry. However under the influence of an applied magnetic field there will be a preferential redistribution of the Co^{2+} among the four types of sites, since their energies will not be equal in the field, and this state is preserved on cooling. Further studies by Bickford, Brownlow and Penoyer (1958) showed that cation vacancies are involved in the rate at which annealing occurs (as might be expected if the Co^{2+} ions migrate at elevated temperatures). By studying the annealing as a function of Co^{2+} concentration they claimed to show also that the formation of Néel type pairs becomes important at higher concentrations, but this conclusion has been criticized by Smit and Wijn (1959) (see also Inoue, Mizuta and Iida 1960). For very low cobalt concentrations magnetic annealing therefore seems to be quite simple and well understood, but in more concentrated ' cobalt ferrites ' there may be several competing processes.

The rather different kind of magnetic annealing effect, which occurs in Fe_3O_4 near its transition at 119°K, has already been discussed in § 2.2(*a*). An apparently similar effect has been observed in non-stoichiometric nickel ferrite by Dwight and Menyuk (1958) and Menyuk and Dwight (1959), who found that by cooling a crystal in a magnetic field below 10°K an additional uniaxial anisotropy was induced, and this depended on the direction of the annealing field. It seems unlikely that electronic ordering is responsible at so low a temperature (the resistivity is greater than 5×10^{13} ohm cm), and they suggested that the effect may in some way be due to excess ions in normally vacant octahedral sites interacting with the other ions.

3.6. *Magnetostriction*

Magnetostriction similar to that observed in ferromagnetic metals (see for example Bozorth 1951) occurs also in ferrimagnetic compounds, but so far no peculiarly different effects have been observed. A summary of the measured magnetostricton coefficients has been given by Bozorth, Tilden and Williams (1955) and by Tsuya (1957). The two most striking results are those for $CoFe_2O_4$ (Bozorth and Walker 1952), which has a very large λ_{100} (-515×10^{-6}), and for Fe_3O_4, which is the only ferrimagnetic compound with positive magnetostriction (Bickford, Pappis and Stull 1956). Some interesting measurements on GdIG have recently been reported by Belov and Pedko (1960). Nearly all the experiments have employed the same method, involving the use of a small strain gauge (Goldman 1947) to measure the change of length for several applied fields (Bozorth and Hamming 1953). An alternative, but rather difficult, method has been reported by Guillaud and Sage (1953), who measured the effect in a cobalt ferrite by means of high order x-ray reflections. The effect of magnetostriction on x-ray diffraction has also been observed by Merz (1960), who used it to study domain structures in Co-Zn ferrite. Other interesting magnetostrictive effects have also been observed in crystals of $Ga_{2-x}Fe_xO_3$, $0{\cdot}6 \leqslant x \leqslant 1{\cdot}4$ (White 1960). These compounds, recently discovered by Remeika (1960), are both ferrimagnetic and piezoelectric, so that mechanical strains can both be induced and detected by either electrical or magnetic means. This suggests a variety of possible experiments and also practical applications.

A very extensive review of the theory of magnetostriction in ferrites has been given by Tsuya (1957, 1958), who has shown that the same mechanisms which account for crystalline anisotropy can contribute to this effect, but that the general position is even more complicated. The only compound for which the theory is again fairly definite is Fe_3O_4 containing a small amount ($\sim 1\%$) of Co^{2+}, for which Slonczewski (1959, 1960) has shown that the main effect comes from the orbital degeneracy of single Co^{2+} ions, as for the anisotropy.

§ 4. Thermal Properties

4.1. *Specific Heat*

Comparatively few specific heat measurements have been reported for ferrimagnetic substances, which is surprising because of the popularity of the method for studying antiferromagnets, and the basic magnetic information which can be obtained in this way. There are at least four magnetic effects which can be studied by means of the specific heat.

(*a*) *Ordering at the Néel point*: As ordering sets in near the Néel point the entropy decreases with temperature, and this corresponds to an additional contribution to the specific heat. In ferromagnetic and antiferromagnetic substances a sharply peaked 'λ type' anomaly is usually observed, and the peak of the anomaly is often taken as the definition of the Curie or Néel point, though this may not always be exactly the same as the temperature at which long range order sets in. If an estimate of the 'non-magnetic' part of the specific heat can be made, either by interpolation from temperatures far removed from the anomaly or by comparison with similar substances which do not show an anomaly, the change of entropy can be calculated by integration of the excess specific heat. This can be compared with the magnetic entropy of $R\ln(2S+1)$ for each type of spin present. An estimate can also be made of the degree of short range order above the Néel point. Such information supplements the normal magnetic data and acts as a check on theoretical models of ordering, but so far only one such experiment seems to have been made specifically on ferrimagnetic compounds (Inoue and Iida 1958, $CuFe_2O_4$). An interesting anomaly in the specific heat has been observed in $NiCr_2O_4$ at 310°K, far above its ferrimagnetic Néel point (~ 80°K) (Lotgering 1956 b), and this has been interpreted as possibly being due to ordering of the type $*\updownarrow$ (see § 3.2), in which one sub-lattice is antiferromagnetic while the other is still paramagnetic (but see also Wojtowicz 1959).

(*b*) *Crystalline phase changes*: A 'λ' type specific heat anomaly may also indicate a change of phase below the Néel point, and this effect has been observed by Millar (1929) in Fe_3O_4 at about 120°K. In this case the ordering mechanism is now well understood (Verwey, Haayman and Romeijn 1947) but the specific heat may be a useful tool for studying phase changes in other substances, such as non-stoichiometric $NiFe_2O_4$ (Menyuk and Dwight 1958) for which the evidence is less direct.

(*c*) *Spin waves*: Probably the most important magnetic specific heat effect is the low temperature contribution from spin waves. For any substance in which the energy of spin waves, $\hbar\omega_k$, is related to the wave vector $\mathbf{k}$ by the dispersion relation

$$\hbar\omega_k = Dk^2 \qquad \ldots\ldots(4.1)$$

as discussed in § 3.2.2, there is a specific heat per unit volume

$$C_v = ck_B\left[\frac{k_B T}{D}\right]^{3/2} \qquad \ldots\ldots(4.2)$$

where k_B is the Boltzmann constant, and c is a constant which depends on the crystal structure and is equal to 0·113 for cubic lattices (see for example van Kranendonk and Van Vleck 1958). The constant D depends on the strength of the exchange interactions and is roughly of order $a^2 k_B T_K$, where a is the mean lattice spacing. This corresponds to a rather small specific heat, which is distinguishable from the Debye lattice specific heat only at low temperatures.

The first experiments to detect this term were reported by Kouvel (1956), who measured the specific heat of Fe_3O_4 between 1·8° and 4°K. In this range the lattice contribution is proportional to T^3, so that the results may be analysed by plotting $C_v/T^{3/2}$ against $T^{3/2}$, as shown in Fig. 10. The evidence for a $T^{3/2}$ term in

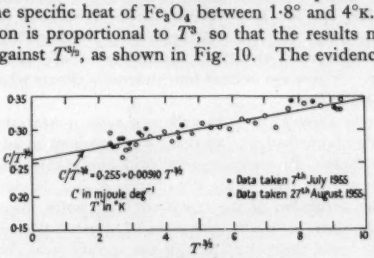


Fig. 10. Low temperature specific heat of Fe_3O_4 showing for the first time a dominant $T^{3/2}$ contribution from spin waves (Kouvel 1956).

the specific heat is clear, and similar experiments have since been carried out on YIG by Edmonds and Petersen (1959, 1960) and Meyer and Harris (1960). The constant D can be related to the various exchange interactions between nearest neighbours, but as in the case of the low temperature magnetization it is probably better to regard it simply as an empirical parameter until there are direct measurements of the interactions. However the results of specific heat measurements can be compared directly with those of other experiments (e.g. neutron scattering (§ 6.2); thermal conductivity (§ 4.2)), even without a detailed analysis of the exchange interactions, since the same constant D appears in the formulae for other effects. It is important to note that Eqn (4.1) is itself an approximation which may be poor for low energy spin waves, i.e. those important for specific heats at very low temperatures. Corrections may arise from demagnetizing effects, even in the absence of an applied field, since there are then domains, and also from anisotropy, both local and overall. If measurements are made in an applied field H_0 there is an extra term on the right-hand side of Eqn (4.1) equal to $\hbar\gamma_e H_0$ where γ_e is the 'effective' gyromagnetic ratio (see § 5.3), and this should lead to an exponential fall-off of the specific heat when $k_B T \ll \hbar\gamma_e H_0$. So far there has been no report of any such measurements.

Since Eqn (4.1) applies to ferro- as well as ferrimagnetics it may seem strange that a $T^{3/2}$ specific heat had not been observed before. The reasons for this is simply that in electrically conducting media there is in addition a large term linear

in T, which would make any analysis very uncertain. At times there have also been doubts whether spin waves can exist at all when there are non-localized electrons (van Kranendonk and Van Vleck 1958), a difficulty which is avoided in almost all ferrimagnetic substances.

(*d*) *Crystal field and exchange splittings*: As discussed in § 3, certain magnetic ions may have low lying energy states in the combined exchange and crystal fields. These will give rise to a Schottky type specific heat anomaly at temperatures where kT is of the order of the splittings, and by fitting the detailed shape of the anomaly it may be possible to deduce values of the exchange interaction constants. Such an analysis has recently been reported by Meyer and Harris (1960), who measured the specific heat of various rare earth iron garnets between 1·4° and 20°K. They found extremely large values (10 to 60 joule mole^{-1}) for all of them except YIG (in which Y^{3+} is diamagnetic), and these can be interpreted in terms of a relatively weak exchange interaction between the iron sub-lattices and the magnetic rare earth ions. In HoIG they observed an additional small contribution below 4°K which may be due to nuclear hyperfine structure. At even lower temperatures this effect should become observable also in other cases.

4.2. *Thermal Conductivity*

The concept of spin waves as thermal excitations having a specific heat and velocity of propagation $v = \partial\omega_k/\partial k$ leads to the prediction that they should also contribute to heat transfer at low temperatures (Sato 1955). This effect has been investigated experimentally by Douthett and Friedberg (1958), who measured the thermal conductivity K of single crystals of three ferrites, $MnFe_2O_4$, $(CoZn)Fe_2O_4$ and $(MnZn)Fe_2O_4$, between 1·4° and 25°K. They found some very unexpected results which indicate that the effect of the spin waves is rather more complex than had been expected, but that at the same time an understanding of this effect may throw light on the important problem of internal relaxation processes.

First of all they found that all three samples had an appreciably *smaller* thermal conductivity than similar non-magnetic substances, so that the spin waves, so far from adding to the conductivity, seem to reduce it. For all three samples the temperature variation of K was similar and roughly proportional to T^2 or T^3, but the absolute values for one of them, $(MnZn)Fe_2O_4$, were about five times larger than for the other two (25 mW cm^{-1} deg^{-1} as against 5 mW cm^{-1} deg^{-1} at 10°K). Secondly they found that the conductivity of the two samples with the lower K could be changed by the application of a magnetic field parallel to the heat flow, while the other showed no effect. In general by applying a field one raises the energy of the spin waves and so decreases their degree of excitation at a given temperature, and one would expect therefore to decrease any conductivity due to the spin waves. Experimentally however an *increase* of K was found, a field of 10 k gauss producing a change of about 20% in both the samples which showed an effect.

They interpreted these results by assuming that the main part of the thermal conductivity in all these samples is due to lattice vibrations as in normal non-magnetic insulators, and that this conductivity is reduced by scattering of the phonons by the spin waves. If the scattering in $(MnZn)Fe_2O_4$ is less, possibly due to a smaller magneto-elastic coupling, changing the number of spin waves would

not affect the conductivity, while for the other two samples the application of a field would reduce the scattering and so increase the conductivity. This would also explain why the conductivity of $(MnZn)Fe_2O_4$ is larger than for the other two. While this explanation is quite reasonable and self consistent, it is clear that a considerable amount of experimental and theoretical work is still called for. In particular, as Douthett and Friedberg have pointed out, measurements are required on compounds which do not have the uncertain composition and randomness in crystal structure of the inverse spinels, since these factors could have an important effect on scattering processes. Also measurements in more than one value of field would help to eliminate the possibility of effects from domain structure, demagnetizing corrections and anisotropy.

4.3. *Magnetocaloric Effects*

When a magnetic substance is suddenly subjected to a change of applied field there is a change of temperature which is related to the change of magnetization. This is called the magnetocaloric effect and it has been studied extensively in ferromagnetic metals by Stoner and Rhodes (1949), Bates and his co-workers, and others (see Bates 1951 a, b, Bates and Marshall 1953). Recently the measurements have been extended to ferrimagnetic substances (Bates and Christoffel 1957). An elegant method of measuring the small changes in temperature particularly suitable for ferrites has recently been devised by Závěta (reported by Valenta 1959). In general at low fields, less than about 500 gauss, the mechanism is very complicated in that it involves irreversible as well as reversible changes of domain wall motion and rotation. The theory of Stoner and Rhodes is able to account for a large part of the observed temperature changes, but at the lowest fields, of the order of 10 to 15 gauss, there are various unexplained effects (see Bates and Christoffel 1957). The interest of measurements of this kind lies in the information regarding the intricate processes of magnetization which can be studied in this way.

For fields large enough to produce technical saturation the position is much simpler and the change of temperature ΔT can be calculated from the thermodynamic relation

$$\Delta T = \frac{-T}{C_H}\left(\frac{\partial M}{\partial T}\right)_H \Delta H, \qquad \ldots\ldots(4.3)$$

where C_H is the specific heat at constant field H, which may generally be taken to be constant and equal to the zero field specific heat C_p or C_v. So far no measurements of this kind seem to have been reported for any ferrimagnetic substances but the effect should be as large as it is in metals, for which changes of the order of 1°K have been observed near the Curie point where $(\partial M/\partial T)_H$ is large (Weiss and Forrer 1926). This formula also predicts a large effect at very low temperatures where the specific heat is small and mainly magnetic. If both M and C_v can be represented by the spin wave theory and $C_H \simeq (C_v)_{H=0}$, then

$$\frac{-T}{C_H}\left(\frac{\partial M}{\partial T}\right)_H = \frac{\frac{3}{2}Ng\mu_B S}{p(2Nk_B S)},$$

where p is a structure dependent constant which is equal to 0·96 for cubic lattices (van Kranendonk and Van Vleck 1958. N.B.—an erratum on page 8 deleting the

solidus in line 1). Therefore

$$\Delta T \simeq 0{\cdot}78\frac{g\mu_B}{k_B}\Delta H$$

which is equal to 1°K for $\Delta H = 10$ k gauss and $g = 2$. In practice this would be reduced by the lattice specific heat, and the quantitative interpretation would also be complicated by the variation of $(\partial M/\partial T)_H$ and C_H with field. However an analysis of measurements of this effect should provide a useful check of the spin wave or other low temperature theories.† They would best be made on insulating ferrimagnets to avoid effects from eddy currents and electronic specific heat.

§ 5. Microwave Properties

5.1. *Dynamic Susceptibility*

As discussed in the introduction the behaviour of ferrimagnetic substances is of particular interest at high frequencies, since for most of them one can observe bulk magnetic properties without interference from eddy current effects. When there are domains present the high frequency behaviour may be very complex (see § 5.9), but when a field $\mathbf{H}_0$ large enough to produce technical saturation is applied the response to a small alternating field $\mathbf{h}$ can be described by a simple susceptibility χ. However due to the fact that the magnetic moment is associated with angular momentum the relationship between the induced alternating moment $\mathbf{m}$ and $\mathbf{h}$ has a tensor form :

$$\mathbf{m} = \boldsymbol{\chi}.\mathbf{h} = \begin{bmatrix} \chi_{11} & j\alpha & 0 \\ -j\alpha & \chi_{22} & 0 \\ 0 & 0 & \chi_{33} \end{bmatrix} . \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} \qquad \ldots\ldots(5.1)$$

where the 3-axis of $\boldsymbol{\chi}$ coincides with $\mathbf{H}_0$. (This formula applies strictly only if crystalline anisotropy is negligible (Artman 1956).) In most cases $\chi_{11} = \chi_{22} = \chi$ and $\chi_{33} \simeq 0$, and both χ and α have real and imaginary components, which depend on the frequency, on H_0, damping terms, the bulk magnetization M_s and the 'effective gyromagnetic ratio' (see § 5.2). Calculated expressions for χ', χ'', α' and α'' in terms of these parameters have been given by Polder (1949), Hogan (1952, 1953, 1956), Artman (1956), Bloembergen (1956) and others, but in practice they are best determined empirically. Methods for doing this have been described by Artman and Tannenwald (1953, 1955), Spencer, LeCraw and Reggia (1956) and Spencer, Ault and LeCraw (1956).

The measured susceptibility depends on the shape of the sample and it is strictly defined only for ellipsoidal shapes, for which it varies with the demagnetizing tensor $\mathbf{N}$. This dependence on shape has given rise to some confusion and much discussion in the literature, which is due to the fact that it is possible to define $\boldsymbol{\chi}$

† *Note added in proof* : A report of extensive magnetocaloric measurements on a polycrystalline sample of YIG has recently been given by Kunzler, Walker and Galt (1960). Allowing for the variation (~30%) of C_H with field, and taking account of the contribution from the lattice, they found good consistency for the data taken over a wide range of temperatures (1·45° to 4·2°K) and fields (3 kG to 18 kG). Comparison with direct calorimetric measurements (see § 4·1(c)) showed a rather considerable discrepancy, but this was claimed to be due to small impurities in the earlier samples.

in one of *two* ways. One can define $\boldsymbol{\chi}$ in terms of the field *inside* the sample $\mathbf{h}_i$, by the relation

$$\mathbf{m} = \boldsymbol{\chi}_i \mathbf{h}_i, \qquad \ldots\ldots(5.2)$$

in which case $\boldsymbol{\chi}_i$ is an intrinsic property of the material independent of sample shape, or else in terms of the *applied* magnetic field $\mathbf{h}_e$ when the 'external' or 'effective' susceptibility $\boldsymbol{\chi}_e$ is given by

$$\mathbf{m} = \boldsymbol{\chi}_e . \mathbf{h}_e, \qquad \ldots\ldots(5.3)$$

which now depends on sample shape. In a cavity perturbation experiment, such as is commonly used to measure $\boldsymbol{\chi}$, the quantity determined is $\boldsymbol{\chi}_e$. For a sample small compared with the wavelength this is related to the intrinsic susceptibility through the equation

$$\mathbf{h}_i = \mathbf{h}_e - \mathbf{N} . \mathbf{m} \qquad \ldots\ldots(5.4)$$

which leads to

$$\boldsymbol{\chi}_i = \boldsymbol{\chi}_e (1 + \mathbf{N} . \boldsymbol{\chi}_i) \qquad \ldots\ldots(5.5)$$

(see also Berk and Lengyel 1955). For a disk-shaped sample with H_0 applied perpendicularly to the disk $N_x = N_y = 0$, $N_z = 4\pi$, and as $(\chi_i)_z \simeq 0$ this gives $(\chi_i)_{\text{disk}} \simeq (\chi_e)_{\text{disk}}$. This fact was pointed out by Rowen and von Aulock (1954) who suggested that using a disk was therefore the obvious way of determining the intrinsic properties experimentally. However this involves using a sample which in two dimensions is large relative to the size of the cavity and the wavelength and, as this can cause complications (see § 5.3(d)), it seems safer to use a sphere and to allow for the difference by Eqn (5.5).

In practice it is often convenient to redefine $\boldsymbol{\chi}$ in terms of circularly polarized components of $\mathbf{h}$, h_+ and h_-, perpendicular to h_z. This transforms Eqn (5.1) to

$$\begin{bmatrix} m_+ \\ m_- \\ m_z \end{bmatrix} = \begin{bmatrix} \chi+\alpha & 0 & 0 \\ 0 & \chi-\alpha & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} h_+ \\ h_- \\ h_z \end{bmatrix} \qquad \ldots\ldots(5.6)$$

assuming $\chi_{33} = 0$. This equation is sometimes further changed into the corresponding relation between the induction $\mathbf{b}$ and $\mathbf{h}$

$$\begin{bmatrix} b_+ \\ b_- \\ b_z \end{bmatrix} = \begin{bmatrix} \mu+K & 0 & 0 \\ 0 & \mu-K & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} h_+ \\ h_- \\ h_z \end{bmatrix} \qquad \ldots\ldots(5.7)$$

where $\mu = 1 + 4\pi\chi$ and $K = 4\pi\alpha$. Typical results of measurements of the real and imaginary parts of $\mu \pm K$ as a function of H_0 for a commercial Mg–Mn ferrite are shown in Fig. 11. The experimental points refer to the 'external' quantities measured on a thin rod sample, and the solid curves show the calculated intrinsic values. It is clear that there is a considerable difference between the two, and it is particularly important to note that when $\mu_e'' + K_e''$ is large $\mu_i'' + K_i''$ is quite small. We shall return to this point in connection with possible size effects in § 5.3(d). The curves also illustrate very clearly that for one sense of circular polarization, defined to be the positive sense, there is a large peak in $\mu_e'' + K_e''$ corresponding

to a resonance absorption at a certain field, while the other sense of polarization has practically no variation. In the next sections we shall discuss in some detail the many factors which determine the value of field at which this ferrimagnetic resonance occurs.

The difference in response to the two senses of circular polarization has important consequences for many practical applications. For example it predicts a difference in the velocities of two oppositely circularly polarized electromagnetic waves, which is equivalent to Faraday rotation of plane polarized waves propagating in the

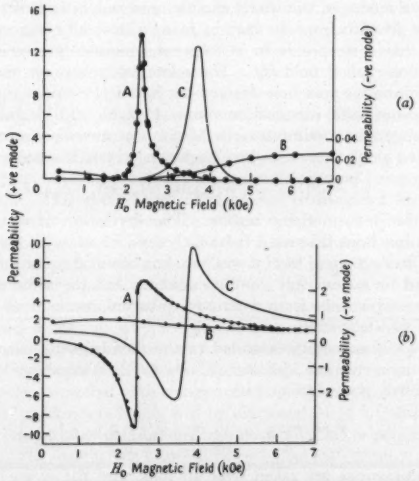


Fig. 11. (*a*) Imaginary part and (*b*) real part of the permeability $\mu_0 \pm \mathbf{K}_0$ for positive (curve A) and negative (curve B) circularly polarized waves for Mg–Mn ferrite at 9·2 Gc/s. Experimental points measured on a thin rod ('external values'); curve C computed for the 'intrinsic' components $\mu_1 + K_1$ (*after* Spencer, Ault and LeCraw 1956).

direction of H_0 (Polder 1949, Roberts 1951). This is a non-reciprocal effect, i.e. one which does not reverse when the direction of propagation is reversed, and it can therefore be used in making microwave devices which permit propagation in one direction but not the reverse (Hogan 1952, 1953, 1956, see also other references under 'Technical Applications').

When an initially plane-polarized wave is propagated perpendicularly to H_0 a similar analysis shows that it will undergo magnetic birefringence and become elliptically polarized. This is the microwave analogue of the optical Cotton–Mouton effect and it has been discussed by Polder (1949), Weiss and Fox (1952) and du Pré (1955). Like the Faraday effect it can be used for non-reciprocal devices, phase shifters and attenuators. The detailed analysis of all such devices

is very complicated as they generally use ' large ' samples for which electromagnetic propagation affects the relation between intrinsic and external properties, and there are also effects from the surrounding microwave structure (see references under ' Technical Applications ').

5.2. *Ferrimagnetic Resonance : Introduction*

Even a small magnetic sample contains about 10^{20} coupled atomic moments, and in principle a very large number of different resonances is possible. However due to the overwhelming effect of the exchange interaction these are coupled to form collective modes of the whole sample, and only a few of these are observed at microwave frequencies. In fact in many cases only one resonance mode is excited and this corresponds to a coherent, uniform precession of the atomic moments in the applied field H_0. The relationship between the frequency $\omega/2\pi$ and H_0 for this mode was first derived by Kittel (1948) for the case of a ferromagnetic medium with identical moments $\mathbf{M}$ each with angular momentum $\mathbf{S}$, for which the gyromagnetic ratio $\gamma = M/S$. For an ellipsoid of such a medium with H_0 applied along the z axis, and neglecting crystalline anisotropy

$$\omega^2 = \gamma^2[H_0 - (N_z - N_x)M_s][H_0 - (N_z - N_y)M_s] \qquad \ldots\ldots(5.8)$$

where $\mathbf{N}$ is the demagnetizing tensor. The derivation of this equation is well known and it has been discussed frequently (see for example Bagguley and Owen 1957). In Kittel's original form it was based on classical equations of motion which are quite valid for extensively coupled systems, but the same result can also be obtained more rigorously from a strictly quantum mechanical approach (Polder 1948, 1949, Van Vleck 1950).

Equation (5.8) is readily extended to the ferrimagnetic case in which not all the magnetic moments are identical (Tsuya 1952, Wangsness 1952) by replacing γ by an ' effective gyromagnetic ratio ' γ_e :

$$\gamma_e = \sum_i \mathbf{M}_i \Big/ \sum_i \mathbf{S}_i = \frac{e}{2mc} \sum_i g_i \mathbf{S}_i \Big/ \sum_i \mathbf{S}_i = g_e\left(\frac{e}{2mc}\right), \qquad \ldots\ldots(5.9)$$

where the summations are taken over all ions, and the g_i are the spectroscopic splitting factors of the individual ions in the usual paramagnetic sense. If the magnetic moment of all the ions arises only from spin the g_i are all 2 and there is no distinction between the ferro- and ferrimagnetic cases. In general however values of g_e appreciably different from 2 are often found, and some of the factors contributing to these differences, both real and apparent, are discussed in the next section. In single crystal specimens the resonance frequency cannot be specified by a single parameter g_e alone, and it is necessary to consider the effect of crystalline anisotropy. This is discussed in § 5.4.

When there is more than one type of sub-lattice a different kind of resonance mode can be excited by inducing a precession in which each sub-lattice responds coherently, but not in phase with the other sub-lattice(s) (Kittel 1951). The energy of such a mode is in general very much higher than that of the uniform mode, since there is a change in the exchange energy as soon as interacting moments are not antiparallel. For a substance with a Néel temperature at a few hundred degrees absolute this type of resonance occurs at infra-red frequencies (Wangsness 1952,

1953, Kaplan and Kittel 1953, Brown and Park 1954), unless a field of the order of the exchange field ($\sim 10^5$ to 10^7 gauss) is applied. Neglecting shape and anisotropy effects, and considering a simple two sub-lattice ferrimagnet for which the exchange energy may be expressed as $n\mathbf{M}_A.\mathbf{M}_B$, the exchange resonance frequency is given by

$$\frac{2mc}{e}\omega_{ex} = g_e H_0 - n g_A g_B \left[\frac{M_B}{g_B} - \frac{M_A}{g_A}\right]. \qquad \ldots\ldots(5.10)$$

In general it is very difficult to obtain either ω or H_0 large enough to observe this resonance, and so far the only experiment which has been reported is one by Foner (1957) who, using pulsed fields of 60 kOe and a frequency of 35 Gc/s, was able to detect absorption in gadolinium iron garnet. For this substance the Gd–Fe interaction is rather weak, and the resonance is therefore more readily accessible. A more promising general method of studying the exchange mode is possible for substances for which $M_B/g_B \simeq M_A/g_A$ at some temperature (Wangsness 1953, 1954). This condition is equivalent to a compensation point in the *angular momentum*, which is proportional to M/g. If, as is frequently the case, $g_A \simeq g_B$, this will occur at temperatures close to (though not at) the *magnetic moment* compensation point mentioned in § 3.2. This special case of resonance is discussed further in § 5.5.

Another important type of resonance which occurs in ferrimagnetics is one which is made possible by the incidental fact that most of them have a low electrical conductivity. Due to this, resonance involves a bulk precession of the magnetic moments throughout the whole sample, and if this is placed in an *in*homogeneous alternating field one of a number of ' non-uniform ' modes of precession may be excited. These are modes in which different parts of the sample precess with different relative phases, so that the instantaneous magnetization is spatially not uniform. This type of resonance will be discussed in § 5.6.

Two other effects connected with resonance, the problems of line width and the response to ' large ' microwave fields, will be discussed in §§ 5.7 and 5.8. Finally in § 5.9 we shall briefly consider some of the dynamic effects which occur in the presence of a d.c. field insufficient to produce technical saturation.

5.3. *Effective g-factor*

In this section we shall consider some of the effects which cause the observed value of g_e to differ from the value 2, and also may cause it to depend on the measuring frequency or temperature. There are at least three intrinsic and six spurious effects which we must consider.

(*a*) *Orbital effects.* The most obvious effect, and one which has been recognized for a long time (Beljers and Polder 1950), is simply the fact that the paramagnetic g-value of some of the ions present is different from 2 due to admixtures of orbital states by the spin–orbit coupling. For example, almost all paramagnetic Ni^{2+} salts have g-values between 2·2 and 2·3 (Bowers and Owen 1955) and it is reasonable to expect a similar value for the Ni^{2+} ions in, say, $NiFe_2O_4$. This would account for the observed value of $g_e = 2{\cdot}26$ (Healy 1952) since the effect of the Fe^{3+} ions cancels in the summations due to the inverse spinel structure of $NiFe_2O_4$. In this case we would expect no variation with temperature, in agreement with experiment. If the effect of the second kind of ion had not cancelled out, a value of g_e different

from the Ni^{2+} ion value would result, and it would vary with temperature since the average value of $\mathbf{S}$ in Eqn (5.9) varies differently for the individual sub-lattices. An extreme case of this is observed near 'compensation points' where $g_e \to 0$ as $\Sigma_i \mathbf{M}_i \to 0$ and $g_e \to \infty$ as $\Sigma_i \mathbf{S}_i \to 0$ (see however § 5.5). However, even at temperatures far removed from such singular points appreciable changes of g_e may occur if the g_i's are very different or the sub-lattices become ordered at a very different rate.

(*b*) *Exchange effects.* The frequency of the uniform precession mode is in general independent of the exchange interaction since the parallel–antiparallel arrangement is preserved at all times. However if there are strong *intra*-sub-lattice exchange interactions the equilibrium state may be a triangular Yafet–Kittel arrangement, and even a coherent precession will then involve the ratios of the various interactions. This case was first considered explicitly by Eskowitz and Wangsness (1957, see also Wangsness 1960), though Tsuya's (1952) quantum mechanical derivation of Eqn (5.9) had already indicated that $\gamma_e = M/S$ is only an approximation which applies for the case of overwhelming inter-sub-lattice coupling. We may similarly expect that the resonance frequency of multi-axis arrangements caused by strong local anisotropy (Fig. 1(*f*)) will also have an anomalous g_e, but in this case the position will be complicated by large spin–orbit–lattice coupling effects on the individual ions. So far there is no direct experimental evidence for these effects as the position regarding multi-axis arrangements is still uncertain (see §§ 3.2 and 3.3), but they will undoubtedly be important in certain cases.

(*c*) *Damping.* So far we have implicitly assumed that damping effects in the equations of motion are small and we have deferred this tricky topic for discussion under 'line width' in § 5.7. However Wangsness (1958) and Kittel (1959) have recently pointed out that very strong damping forces on some of the ions of a ferrimagnet may produce a shift in the resonance frequency, without contributing excessively to the total resonance line width. This case occurs if the effect of the damping is stronger than that of the exchange interaction, so that the gyroscopic effect of one (or more) sub-lattice(s) is effectively quenched. The general case of this has been considered by Wangsness, and it has been applied explicitly by Kittel to account for the observed g's in several rare earth iron garnets, which have values lying between 1 and 2 and which vary with temperature (Rodrigue *et al.* 1958, Rodrigue, Meyer and Jones 1960).

We can illustrate the effect simply by a two sub-lattice example in which the A sub-lattice behaves normally, with negligible damping, while the B sub-lattice is so strongly damped that it responds instantaneously to any change of field acting on it. The classical equations of motion are then

$$\frac{d\mathbf{M}_A}{dt} = \gamma_A(\mathbf{M}_A \times \mathbf{H}_A), \qquad \ldots\ldots(5.11a)$$

$$\mathbf{M}_B = |M_B| \frac{\mathbf{H}_B}{|H_B|}, \qquad \ldots\ldots(5.11b)$$

where the effective fields on the two sub-lattices are

$$\mathbf{H}_A = \mathbf{H}_0 + n\mathbf{M}_B \qquad \ldots\ldots(5.12a)$$

$$\mathbf{H}_B = \mathbf{H}_0 + n\mathbf{M}_A, \qquad \ldots\ldots(5.12b)$$

$\mathbf{H}_0$ is the applied field and n the (A–B) exchange constant. Demagnetizing, anisotropy and intra-sub-lattice coupling effects are neglected. Substituting, Eqn (5.11*a*) becomes

$$\frac{d\mathbf{M}_A}{dt} = \gamma_A\left\{\mathbf{M}_A \times \left[\mathbf{H}_0 + n\frac{|M_B|}{|H_B|}(n\mathbf{M}_A + \mathbf{H}_0)\right]\right\}$$

$$= \gamma_A\left(1 + \frac{n|M_B|}{|H_B|}\right)(\mathbf{M}_A \times \mathbf{H}_0)$$

$$\simeq \gamma_A\left(1 - \frac{|M_B|}{|M_A|}\right)(\mathbf{M}_A \times \mathbf{H}_0),$$

since $\mathbf{H}_0 \ll n\mathbf{M}_A$ and $n < 0$.

This is the usual gyroscopic equation of motion for the A sub-lattice, except that γ_A is replaced by

$$\gamma_e = \gamma_A\left[\frac{|M_A| - |M_B|}{|M_A|}\right] = \frac{M_s}{S_A}. \qquad \ldots\ldots(5.13)$$

The highly damped B sub-lattice thus contributes to the total magnetic moment M_s, but not to the total angular momentum. As M_B is in general not proportional to M_A, γ_e will change with temperature, in a way which can be calculated from a knowledge of the sub-lattice magnetizations and the g-value of the undamped lattice.† At sufficiently low temperatures the effect may be expected to disappear, since in general relaxation processes become slower and the effect of exchange will then restore γ_e to its usual value (Eqn (5.9)).

So far direct experimental evidence has only been found for the fully damped case (Eqn (5.13)) (Kittel 1959), since for the rare earth garnets line width and anisotropy effects complicate low temperature measurements of γ_e. However the effect of damping, either total or partial, may be of importance quite generally for all compounds which contain ions which are strongly bound to the lattice through incomplete quenching of the orbital angular momentum (see § 3.4). It may therefore often be associated with ions which also make a large contribution to the anisotropy (White 1959 b).

(*d*) *Size effect*. It has long been recognized (Yager *et al.* 1950, Beljers and Polder 1950, Yager, Merrit and Guillaud 1951, Miles 1954, Artman 1957 b, Tompkins and Spencer 1957) that the physical size of the sample relative to the wavelength of the microwaves can have an important effect on the frequency of resonance, but until recently there has been little agreement about the exact cause of the effect. It used to be thought that the effect was simply a form of ' dimensional resonance ' due to standing waves inside the sample, resulting from a reduction of wavelength by the large value of μ near resonance. However this view fails to recognize the difference between ' external ' and ' intrinsic ' permeability, and

† *Note added in proof* : Eqn (5.13) has recently been applied by Miyadai *et al.* (1960) to explain the observed, temperature dependent g-value of europium iron garnet. The apparent agreement was very good, but Wolf (1960) has pointed out that an equation identical to (5.13) can also be derived for the special case of europium compounds without invoking rapid relaxation. This is because the exchange field, which induces a magnetic moment into the otherwise non-magnetic $\mathbf{J} = 0$ ground state of the Eu^{2+} ion resulting from a dominant spin–orbit coupling, does not produce any corresponding angular momentum. Van Vleck (to be published) has shown that this is a special case of a situation which may arise more generally, in that the angular momentum can also be effectively quenched by strong crystalline electric fields.

as we have seen in § 5.1 this can be considerable. In particular, the intrinsic permeability, which determines the wavelength, is generally quite small when the external permeability shows resonance, so that a dimensional resonance will occur only for relatively very large samples. (The case of a thin disk raises a special problem, which however is generally of no practical importance.)

More recently the size effect has been reconsidered by Hurd (1958) and by Mercerau (1959). Hurd calculated the perturbation of a finite size ellipsoid on the microwave field outside the body, while Mercerau solved the same problem from the point of view of the electromagnetic fields inside the body. Both methods lead essentially to the same answer, which in the case of a sphere gives for the resonance frequency

$$\omega = \gamma_{R=0} H_0 \left[1 - \frac{8\pi M_s}{9H_0} \left(1 + \frac{\epsilon}{5} \right) k^2 R^2 \right], \qquad \ldots\ldots (5.14)$$

where k is the wave number for the microwave field outside the sphere ($k = 2\pi/\lambda$), ϵ is the dielectric constant and R is the radius. The form of this variation is in agreement with the empirical law found by Miles (1954), and it is in quantitative agreement with experiments by Solt and Carson (unpublished, quoted by Mercerau 1959).

From a practical point of view Eqn (5.14) shows that the shift in the resonance frequency is appreciable for spheres larger than about $\lambda/30$. For example taking $4\pi M_s = 2000$ gauss, $\epsilon = 10$, $\lambda = 3$ cm, $H_0 \simeq 3000$ Oe and $R = 1$ mm, the shift $\Delta\omega/\omega$ is 2%. Thus, rather small samples must be used to get accurate effective g-factors ; the problem of making these is closely related to the next factor which we consider.

(*e*) *Sample shape.* Kittel's resonance condition Eqn (5.8) shows that any deviation of the sample shape from that assumed in the calculation may lead to a considerable error in the deduced value of g_e. For example, if instead of a true sphere we use an ellipsoid with an axial ratio 10:9 then with the values of M_s and H_0 above $\Delta\omega/\omega = 2\%$. In a polycrystalline sample this effect can be detected by varying the orientation of H_0, but in a single crystal such an effect will appear as a spurious anisotropy.

The method of making small spheres which is usually employed is that of tumbling the specimen inside a cylindrical abrasive drum by means of a jet of air (Bond 1951, 1954, Senio and Tucker 1953), but Dillon (1958 d) has pointed out that this method is really satisfactory only if there is no anisotropy in the hardness. This is often not the case, especially in non-cubic crystals. He showed that much better spheres could be obtained by grinding a roughly shaped sample between two rotating tubes, as described by Bond (1954), and he was able to obtain true spheres as small as 0·25 mm radius. In this way it is also possible to produce a very smooth surface, and this is important for reducing line widths (see § 5.7).

(*f*) *Non-uniform excitation.* In the derivation of the resonance condition (5.8) it is assumed that the microwave field is effectively uniform over the volume of the sample, and this is generally achieved experimentally by positioning the sample at a suitable point in the microwave cavity. However, Dillon (1958 d) has pointed out that the dielectric inhomogeneity introduced by most types of sample support may cause a significant distortion of the ideal field configuration, and he devised a special mounting technique to overcome this effect. If the radio-frequency field

is not uniform there may be excitation of magnetostatic modes (see § 5.5) which may make the determination of the field for uniform resonance ambiguous. Non-uniform modes may also be excited when there exist surface irregularities in the sample, and if these modes are almost degenerate in energy with the uniform mode an appreciable shift in the resonance field may result (Fletcher and Solt 1959). This may be avoided by careful polishing and mounting of the sample.

(*g*) *Wall effect*. It is well known that the proximity of the sample to the wall or end of the microwave cavity can shift the resonance frequency (see e.g. Miles 1954). This has been investigated experimentally by Spencer and LeCraw (1955) who showed that there is an appreciable effect for distances less than about 2 sphere diameters between the sample and the wall, which arises from eddy currents induced by the alternating magnetic moment. They pointed out that in rectangular cavities this effect will be further complicated by the fact that a linearly polarized radio-frequency field in one direction induces a moment at right angles, due to the off diagonal components of χ, and they advocated the use of square or cylindrical cavities. Further investigations of these effects have been reported by Snieder (1958), Hauser and Brown (1959) and Nash and Standley (1960). Probably the safest way of avoiding trouble is to use a full wavelength transverse electric mode cavity with a very small sample placed at the centre.

(*h*) *Porosity*. In polycrystalline samples, on which many of the earlier measurements were made, it was frequently observed that the effective *g*-factor varies with measuring frequency. Okamura, Torizuka and Kojima (1952) showed that this variation could be removed by rewriting the resonance conditions for a sphere in the form

$$\omega = g_e \frac{e}{2mc}(H_0 + H_i),$$

where H_i is an ' apparent internal field '. Kojima (1954) and McGuire (1955 b), noting the effect of density on the effective *g*-factor in polycrystalline samples, suggested that this internal field must, in part at least, be dependent on the porosity of the sample. This was subsequently confirmed by Schlömann (1956), who treated the effect of porosity theoretically and showed that to a first approximation

$$H_i = \frac{4\pi M_s}{3}\frac{v}{V},$$

where V is the volume of the sample and v is the volume of all pores. For a highly sintered polycrystalline sample v/V may be of the order of 0·05, so that H_i will be about 30 Oe for $4\pi M_s = 2000$ gauss. At a frequency of 10 Gc/s and a true *g*-value of 2 the error in g_e is then about 1%. For a poorly sintered sample the error can be considerably larger.

(*i*) *Anisotropy*. The effect of crystalline anisotropy will be treated in the next section and we note here only the fact that even in polycrystalline samples the field at the peak of the absorption curve $\mu'' + K''$ does not correspond exactly to the true value of g_e if the individual crystallites are appreciably anisotropic. The shift can again be expressed in terms of a contribution to the internal field H_i, which in this case is directly proportional to the anisotropy. This varies with temperature so that g_e will apparently depend on both frequency and temperature. Both this effect and that of porosity are eliminated by using single crystals.

Conclusion. From the above discussion it can be seen that any determination of the effective g-factor must be interpreted very carefully. The dates of the references show that the importance of some of the factors has only been appreciated recently and caution must therefore be exercised in comparing some of the early values quoted in the literature. Tables of collected experimental results may be found in Bagguley and Owen (1957) and Smit and Wijn (1959).

5.4. *Effect of Anisotropy on Resonance*

(*a*) *Single crystals.* The methods used for calculating the effect of crystalline anisotropy on the uniform resonance mode of a ferrimagnetic system have been reviewed by Bloembergen (1956). The most general of these derives the classical equation of motion of the magnetization from an expression for the total free energy F of the system. As discussed in § 3.4, this describes the crystalline anisotropy energy as a function of the orientation of the total magnetic moment $\mathbf{M}$ of all sub-lattices, and it also includes terms from the interaction of $\mathbf{M}$ with $\mathbf{H}$, and demagnetizing effects. The use of a single free energy equation makes the implicit assumption that the inter-sub-lattice exchange coupling is very much stronger than the effect of the anisotropy or the applied field, so that the entire system precesses in phase, but this is generally a good approximation (see however § 5.5 ; also Fletcher and Solt 1960). Expressing F in terms of polar angles, θ and ϕ, relative to the crystal axes, the resonance frequency is then (Smit and Beljers 1955, Suhl 1955)

$$\omega = \frac{\gamma_e}{M_s \sin\theta_0}\left[\frac{\partial^2 F}{\partial\theta^2}\frac{\partial^2 F}{\partial\phi^2} - \left(\frac{\partial^2 F}{\partial\theta\,\partial\phi}\right)^2\right]^{1/2}, \qquad \ldots\ldots(5.15)$$

where the derivatives of F are evaluated for values θ_0 and ϕ_0 corresponding to the static conditions for stable equilibrium

$$\frac{\partial F}{\partial\theta} = 0\,; \quad \frac{\partial F}{\partial\phi} = 0\,; \quad \frac{\partial^2 F}{\partial\theta^2}\cdot\frac{\partial^2 F}{\partial\phi^2} > \left(\frac{\partial^2 F}{\partial\theta\,\partial\phi}\right)^2.$$

Difficulties at $\theta_0 = 0$ can be avoided by redefining θ and ϕ with respect to other axes.†

In general this procedure is not very convenient and it can be simplified appreciably by considering only cases in which the equilibrium orientation of $\mathbf{M}$ coincides with one of the principal axes of the crystal. The effect of the anisotropy can then be represented as an effective anisotropy field $\mathbf{H}^a$, and the resonance condition Eqn (5.8) rewritten as

$$\left(\frac{\omega}{\gamma_e}\right)^2 = [H_0 + (N_x - N_z + N_x^a)M_s][H_0 + (N_y - N_z + N_y^a)M_s] \qquad \ldots\ldots(5.16)$$

(Kittel 1948, 1949), where $N_x^a M_s = H_x^a$ and $N_y^a M_s = H_y^a$ and $\mathbf{H}^a$ depends on F and the angle of H_0 relative to the crystal. For a crystal with cubic symmetry for which

† It must be noted that Eqn (5.15) applies only if γ_e as defined by Eqn (5.9) is a scalar quantity. If the effects of local crystalline anisotropy are large the individual sub-lattice g-values may be highly anisotropic, and the resonance equations would then require very careful re-examination. So far this case has not been treated in the literature.

F can be expressed in the form of Eqn (3.7) this leads to resonance conditions for $\mathbf{H}_0$ along each of the three principal directions of the form :

$$[100] \qquad \omega = \gamma_e(H_0 + 2K_1/M_s)$$

$$[111] \qquad \omega = \gamma_e(H_0 - 4K_1/3M_s - 4K_2/9M_s)$$

$$[110] \qquad \omega = \gamma_e[(H_0 - 2K_1/M_s)(H_0 + K_1/M_s + K_2/2M_s)]^{1/2}$$

(Kittel 1948, 1949, Van Vleck 1950, Bickford 1950, Dillon, Geschwind and Jaccarino 1955). When the anisotropy is small Eqn (5.16) applies approximately also for orientations of $\mathbf{H}_0$ other than in the principal directions, and it is usual to

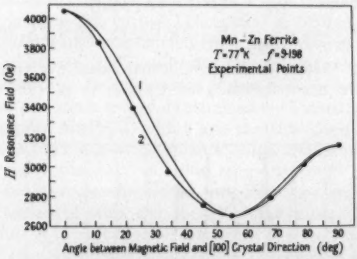


Fig. 12. Typical variation of resonance field with angle in the (110) plane for a substance with negative cubic (K_1, $K_2 < 0$) anisotropy. Curve 1 represents Eqn (5.17) fitted at turning points and curve 2 is corrected to allow for non-alignment of $\mathbf{M}$ and $\mathbf{H}$ (*after* Artman 1956).

consider the case of H_0 in a (110) plane, since measurements as a function of angle then include also the three principal directions. In this case N_x^a and N_y^a are given by (Kittel 1948, 1949, Van Vleck 1950, Bickford 1950, Dillon, Geschwind and Jaccarino 1955)

$$N_x^a = \frac{K_1}{M_s^2}(2 - \sin^2\theta - 3\sin^2 2\theta) + \frac{K_2}{2M_s^2}\sin^2\theta(6\cos^4\theta - 11\sin^2\theta\cos^2\theta + \sin^4\theta),$$

$$N_y^a = \frac{K_1}{M_s^2}(2 - 4\sin^2\theta - \tfrac{3}{4}\sin^2 2\theta) - \frac{K_2}{2M_s^2}\sin^2\theta\cos^2\theta(3\sin^2\theta + 2), \qquad \ldots\ldots(5.17)$$

where θ is the angle between H_0 and the [001] axis. The form of this variation is illustrated in Fig. 12, curve 1.

The error involved in using this kind of expression for an arbitrary orientation of H_0 (arising from the fact that $\mathbf{M}$ and $\mathbf{H}_0$ are then not parallel) has been discussed by Artman (1956, 1957 a), who has given general nomograms of ω/γ against H_0 for various orientations, both for the cases of positive and negative first-order cubic anisotropy. If the anisotropy is not too big, a small correction term can be applied analytically in Eqn (5.17). The improved agreement which can be obtained is shown in Fig. 12, curve 2.

In uniaxial crystals the anisotropy is generally very much larger and it is necessary to use the exact form (5.15), which however is then quite simple to use due to the higher symmetry. If the anisotropy can be expanded in the form of Eqn (3.8), and if, as is often the case, $K_1' \gg K_2'$, etc., the resonance conditions become: For $K_1' > 0$:

$$H_0 \| \text{ axis } \omega/\gamma_e = H_0 + 2K_1'/M_s$$

$$H_0 \perp \text{ axis } \omega/\gamma_e = [H_0(H_0 - 2K_1'/M_s)]^{1/2}; \quad H_0 > 2K_1'/M_s$$

(Smit and Beljers 1955, Weiss and Anderson 1955).
For $K_1' < 0$:

$$H_0 \perp \text{ axis } (\omega/\gamma_e)^2 = H_0(H_0 + |2K_1'/M_s|)$$

$$H_0 \| \text{ axis } \omega/\gamma_e = (H_0 - 2K_1'/M_s); \quad H_0 > |2K_1'/M_s|$$

(Jarrett and Waring 1958). More complicated cases involving high order anisotropy constants have been discussed by Casimir *et al.* (1959) and Belson and Kriessman (1959 b).

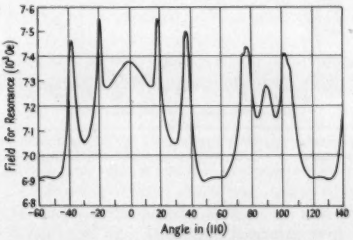


Fig. 13. Variation of resonance field with angle in the (110) plane for YIG containing 0·019% terbium, measured at 20 470 Mc/s and 1·5°K (*after* Dillon and Nielsen 1960 b).

All the above equations hold only for applied fields large enough to remove any domain structure. In lower fields multiple resonances are often observed (see § 5.9.1), and for a quantitative estimate of the anisotropy such effects are best avoided by using a larger field and higher measuring frequencies. In cubic crystals non-saturation effects will generally be important only at low temperatures where the anisotropy becomes large (e.g. Tannenwald 1955 a, b), when special care is needed in any case to ensure that the expansion of F is adequately convergent.

A very striking case in which a low temperature cubic anisotropy cannot be fitted by the usual expansions has recently been reported by Dillon and Nielsen (1959). They found that yttrium iron garnet containing small impurities of rare earths (e.g. 0·02%Tb) has a resonance which is normal at room temperature but which shows a most anomalous variation of resonance field with angle at low temperatures ($< 20°$K) (see Fig. 13). It is clear that this cannot be fitted by the usual type of angular function (Eqn (5.17)). Kittel (1960) has suggested that the sharp peaks arise from a discontinuity in the variation of energy with angle, caused by

the accidental crossing of atomic energy levels of the rare earth ions as the orientation of the internal exchange field acting on them varies. In the limiting case of $T = 0°\text{K}$ this corresponds to a δ-function in $\partial^2 F/\partial\theta^2$, which at finite temperatures becomes a sharp peak in H_{res} against θ, with a width proportional to T, as found experimentally. Wolf (1959) has pointed out that similar sharp peaks can also arise from normal crystal field effects on the individual rare earth ions, whose energy in the exchange field depends on direction because of a kind of ' anisotropic g-factor '. For this mechanism the maxima are closely related to the principal axes of the crystal, as found for various rare earths other than Tb^{3+} (Dillon and Nielsen 1960 a, b), and in some cases both effects may be important. These experiments illustrate again that extreme purity is necessary in certain cases, and they also explain other anomalous results which had previously been found on supposedly ' pure ' yttrium iron garnet (Dillon 1957 a, 1958 c).

A table of anisotropy constants derived from microwave experiments has been given by Bagguley and Owen (1957) and some more recent values may be found in the references cited above. In a few cases it has been possible to compare the results of microwave and static torque measurements (Bozorth *et al.* 1955). For a sample of relatively pure $MnFe_2O_4$ with a low electrical conductivity good agreement was found, but for a non-stoichiometric, high conductivity nickel ferrite a considerable discrepancy was observed at low temperatures. This was interpreted in terms of a relaxation process connected with Fe^{2+} ions, the characteristic time of which is long compared with the period of the microwaves. Under these conditions the essentially thermodynamic treatment of anisotropy must be modified to include only the effect of those systems which are effectively in equilibrium. A similar discrepancy may be expected for the rare earth iron garnets on the basis of Kittel's explanation of the g-values (see § 5.3(c)).†

(b) *Polycrystalline samples.* The effect of anisotropy in polycrystalline samples has been considered frequently (Van Vleck 1950, Kittel and Abrahams 1953, Miles 1954, Standley and Stevens 1956, Schlömann 1956, 1958 a, b, Geschwind and Clogston 1957, Schlömann and Jones 1959), and a good review of the recent work has been given by Schlömann (1959 a).

If the anisotropy is small compared with H_0 but large compared with the internal demagnetizing fields ($\sim 4\pi M_s$), the main effect is a broadening of the resonance line, which may be thought of as a superposition of the resonance of the individual crystallites. To a first approximation the centre of gravity of the line coincides with that predicted in absence of any anisotropy, but as the line shape is generally asymmetric the peak no longer occurs at the same field. The shift may be expressed as an internal field, which for the case of first-order cubic anisotropy (K_1) has been estimated variously as $0{\cdot}2\,K_1/M_s$ (Miles 1954) and $0{\cdot}5\,K_1/M_s$ (Schlömann 1956). The width of the line is about $|K_1|/M_s$.

In many cases the ' independent crystallite ' analysis is not strictly applicable, as has been pointed out by Geschwind and Clogston (1957) and elaborated by

† *Note added in proof* : An interesting anomalous anisotropy effect has recently been observed by Pearson and Teale (1960). Making measurements on crystals of ytterbium iron garnet at two different microwave frequencies, as well as by the static torque method, they found agreement between the three sets of data for certain orientations of the magnetic field (along the cubic [001] axes), but big discrepancies for other orientations (maximum for H_0 along [111]). The origin of this effect is not understood, but it was suggested that it may be connected with the anisotropic coupling between the Yb ions and the Fe sub-lattices, due to local crystalline electric fields.

Schlömann (1958 a). This is because K/M_s is generally smaller than, or comparable with, $4\pi M_s$, and long-range magnetic dipolar interactions then couple the entire assembly into a common, uniform resonance mode. The main effect of this is to narrow the line width by a factor of order $4\pi M_s(K/M_s)^{-1}$, but the peak is still somewhat shifted from the resonance field for zero anisotropy (Clogston 1958).

More complicated effects occur when the anisotropy is large or comparable with both H_0 and $4\pi M_s$. Schlömann and Zeender (1958) have shown that in this case the absorption curve may have subsidiary maxima at fields above and below the normal resonance, the additional peaks becoming well resolved if $|K_1|/M_s$

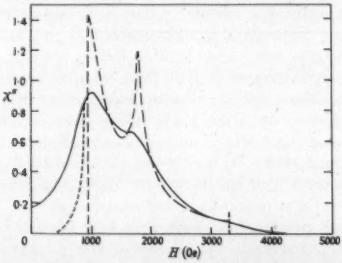


Fig. 14. Typical absorption–field curve, for a polycrystalline sample in which the anisotropy is comparable with $4\pi M_s$. The full line shows the experimental variation for $NiFe_{1.3}Al_{0.7}O_4$ measured at 9 Gc/s ; the broken line the calculated variation. The multiple peaks are due to the anisotropy. Note especially that the middle peak which corresponds approximately to the unperturbed resonance is smaller than the peak at the lower field (*after* Schlömann 1959 a).

exceeds H_0. They point out that this situation will frequently arise in the vicinity of magnetic moment compensation points, since in general K_1 does not pass through zero with M_s, and they suggest that some of the subsidiary compensation point resonances which have been observed (see § 5.5) may be due to this cause (see also Zeender and Schlömann 1960). The detailed theory of this effect has been worked out by Schlömann (1958 b) for the case of cubic symmetry with $K_1 < 0$ and K_2, K_3, etc. small, and compared with experiments on $NiAl_{0.7}Fe_{1.3}O_4$. A typical result is shown in Fig. 14.

Even more striking results are obtained for polycrystalline samples of substances which have axial crystalline symmetry, since for these the anisotropy is in general very large. Three maxima in the absorption curve are then clearly resolved, and in some cases it is possible to deduce the anisotropy constants K_1', K_2', K_3' and K_4' from an analysis of the curve (Schlömann and Jones 1959). In general however the use of single crystals is preferable in all fundamental investigations.

5.5. *Resonance near Compensation Points*

In the discussion of the resonance effects which we have considered so far the characteristic sub-lattice structure of ferrimagnets has been in evidence only indirectly, because we made the generally valid assumption that the effect of the exchange interaction is so overwhelming that the entire sample resonates in one

coherent mode. However this is not always true, and in particular the approximation breaks down at temperatures near a compensation point of either the magnetic moment or the angular momentum. Resonance under these conditions has been discussed extensively by Wangsness (1953, 1954, 1955), Dreyfus (1955 a, b) and Paulevé and Dreyfus (1956). A review of the theory with particular application to the interpretation of some very detailed experiments on single crystals of gadolinium iron garnet has recently been given by Geschwind and Walker (1959), whose general discussion we shall follow.

It turns out that though the detailed behaviour close to compensation points is rather complicated (because of secondary effects arising from sub-lattice susceptibilities, anisotropy and the formation of unusual domains) the general variation of resonance field with temperature can be understood in terms of a simple two sub-lattice model. Using the classical equations of motion of two coupled sub-lattices we find

$$\frac{d\mathbf{M}_1}{dt} = \gamma_1[\mathbf{M}_1 \times (H_0 + H_1^a - n\mathbf{M}_2)]$$

$$\frac{d\mathbf{M}_2}{dt} = \gamma_2[\mathbf{M}_2 \times (H_0 + H_2^a - n\mathbf{M}_1)],$$

where H_0 is the applied field, n the inter-sub-lattice molecular field constant (as in Eqn (5.12)) and H_1^a and H_2^a the effective anisotropy fields of each sub-lattice (as in Eqn (5.16)). For simplicity, demagnetizing effects are neglected and the theory therefore applies strictly only to spherical samples. These equations have solutions for resonance of the form

$$(\omega/\gamma_1 + H_1^a + nM_2 + H_0)(\omega/\gamma_2 - H_2^a - nM_1 + H_0) + n^2 M_1 M_2 = 0. \quad \ldots\ldots (5.18)$$

This is a hyperbola which in general has four roots of H_0 in any given frequency $|\omega/2\pi|$, but two can be eliminated on physical grounds (Dreyfus 1955 b, Paulevé, Dreyfus and Soutif 1959). In this equation the signs have been so chosen that all quantities denote absolute magnitudes except γ_1 and γ_2 which are intrinsically negative for electronic moments. A solution for positive ω then corresponds to the normal type of Lamor precession (anticlockwise when viewed along H_0), and this is the 'ferromagnetic' mode discussed in the previous sections. In this mode M_1 and M_2 normally remain antiparallel as they precess, and the resonance can be described by an effective g-factor (§ 5.3) and a total anisotropy field (§ 5.4). However when $M_1 \simeq M_2$ the solution of Eqn (5.18) shows that the resonance field now depends also on the exchange constant n, corresponding to a flexing of the bond between the sub-lattices. It then becomes very sensitive to small changes in M_1 and M_2 and so varies rapidly with temperature. At the point at which $M_1 = M_2$ there is a sudden discontinuity (see Fig. 15, curve 1) due to the reversal of M_1 and M_2, the larger always being in the direction of H_0.

In order to calculate the exact form of the variation of resonance field with temperature it is necessary to allow for the fact that M_1 and M_2 may themselves depend on H_0 through the differential susceptibility, so that even at the magnetic moment compensation point, which is defined as the cancellation of *spontaneous* moments, M_1 may not be exactly equal to M_2. Although this difference is very

small its effect is magnified by being multiplied by the molecular field nM, which is large relative to H_0. Similarly, the effect of anisotropy is enhanced by the exchange interaction in cross product terms of the kind $nM_1H_2^a$. Thus the anisotropy which in GdIG produces an angular variation in the resonance field of a few hundred gauss at most temperatures, gives rise to a variation of about 5000 gauss at T_c (compare curves 1 and 2 in Fig. 15). In an accurate analysis this effect is further complicated by the fact that H_1^a and H_2^a also depend on temperature, but so far this refinement has not been taken into account.

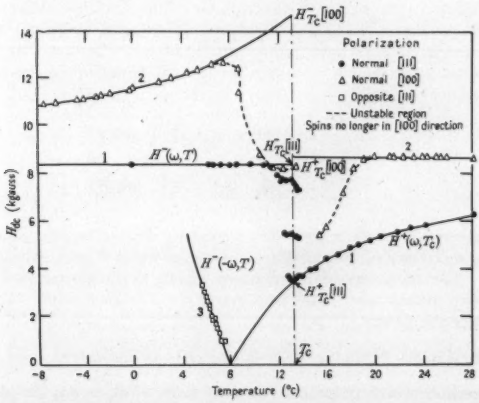


Fig. 15. Variation of resonance field with temperature in GdIG in the region of the magnetic compensation temperature, $T_c = 13°c$. Results are for a single crystal sphere at a frequency of 24 Gc/s. The opposite polarization mode ($-\omega$) was observed in the easy direction [111] between 5·5° and 7°c (curve 3). H^+ and H^- refer to measurements above and below T_c (*after* Geschwind and Walker 1959).

In polycrystalline samples the effect of the enhanced anisotropy is to broaden the resonance beyond the limit of detection in the immediate vicinity of the compensation point, but measurements can generally be made over a sufficient temperature range to illustrate the general behaviour predicted by Eqn (5.18) (see for example van Wieringen 1953, Paulevé 1957 a, b, 1958, Paulevé, Dreyfus and Soutif 1959, Calhoun, Overmeyer and Smith 1957).

In addition to the solution for positive ω Eqn (5.18) has a second root for the inverse sense of polarization. This is the ' exchange mode ' in which M_1 and M_2 precess slightly inclined to one another, as in the normal modes of antiferromagnetic resonance (Keffer and Kittel 1952, Keffer, Kaplan and Yafet 1953). The physical relationship between the different modes has been discussed by Geschwind and Walker (1959), who point out that the ferrimagnetic exchange mode is in fact exactly equivalent to the normal mode of resonance, but corresponding to a negative applied field. This may be seen from Eqn (5.18) by reversing

the signs of ω and H_0 and then interchanging the subscripts 1 and 2, corresponding to the change in H_0. Thus curve 3 in Fig. 15 is the mirror image of the extrapolation of curve 1. As in the case of the normal mode, the exact resonance condition is very sensitive to small changes in M_1, M_2, H_1^a and H_2^a, and H_0 therefore varies rapidly with temperature, but in this case even outside the immediate region of the compensation point. For normal microwave frequencies and magnetic fields (e.g. 24 Gc/s and 10 k gauss) the exchange mode can generally only be observed over a very small temperature region, where $\gamma_1 M_2 \simeq \gamma_2 M_1$, as has been indicated by Eqn (5.10).

The chief information which can be obtained from an analysis of the resonances near a compensation point is the value of the exchange parameter n. In general the value obtained will be rather more accurate than that derived from the susceptibility or the magnetization not only because resonance methods are intrinsically more precise, but also because they are independent of *intra*-sub-lattice interactions (provided these are reasonably small). The other constants which may be deduced are $H_1^a + H_2^a$ and $\gamma_1 - \gamma_2$, though the latter is probably determined more accurately from paramagnetic resonance of the individual ions.

There are several experimental difficulties in observing compensation point resonances: compared with ordinary ferrimagnetic resonance, the intensity is very low, the exchange mode being roughly proportional to $(\gamma_1 - \gamma_2)^2$ (Kaplan and Kittel 1953) while the normal mode depends on $\mathbf{M}_1 + \mathbf{M}_2$. (Very close to T_c the position is more complicated and a resonance can occur even if $\mathbf{M}_1 = -\mathbf{M}_2$ and $\gamma_1 = \gamma_2$, provided the anisotropy is not zero; cf. antiferromagnetic resonance, Keffer and Kittel 1952.) Moreover line widths tend to become large in this region and this reduces the resonance signal still further. It is not quite clear to what extent the increase in line width is an intrinsic effect associated with the compensation (Calhoun, Overmeyer and Smith 1957) or merely the result of small inhomogeneities in the sample (either of composition or temperature) which are accentuated by the critical dependence of H_0 on M_1 and M_2. In polycrystalline samples there is the additional contribution to ΔH from the anomalous anisotropy. This can also give rise to completely spurious secondary peaks, as mentioned in § 5.4, which can be distinguished unambiguously from the exchange mode only by using circularly polarized microwaves, instead of the normal linear polarization which can excite either sense of precession (see for example Zeender and Schlömann 1960). In fact the first report of an exchange resonance (Wangsness 1955, McGuire 1955 a) has been criticized by Paulevé and Dreyfus (1956), and by Schlömann (1959 a), who drew attention to the fact that this ambiguity had not been considered. The difficulty can be avoided by using single crystals, which in any case are necessary for a complete analysis.

However, even with a single crystal one difficulty may be encountered from anisotropy at temperatures very close to T_c. This arises from the formation of ' pseudo-antiferromagnetic domains ' (Geschwind and Walker 1959), which occur because in certain parts of the sample it may be energetically more favourable for M_1 and M_2 to point along one of the easy axes which is *not* parallel to H_0. However this complication arises only within a very small temperature range ($\pm 1°$K in the sample used by Geschwind and Walker) and therefore it does not really confuse the general behaviour of the ordinary resonance modes.

5.6. *Non-uniform Resonance Modes*

A rather different kind of resonance which has been observed in ferrimagnetic crystals occurs when a sample in an external field is subjected to an inhomogeneous microwave field (White and Solt 1956, Dillon 1956, 1960). Under this condition absorption may occur at more than one applied magnetic field for a given frequency, the intensity and position of the resonances depending on the symmetry of the radio-frequency field at the sample. This effect was explained semi-quantitatively by White and Solt (1956) together with Mercerau and Feynman (1956), who showed that the resonances correspond to spatially non-uniform precession modes in which the magnetization in different parts of the sample responds to the driving

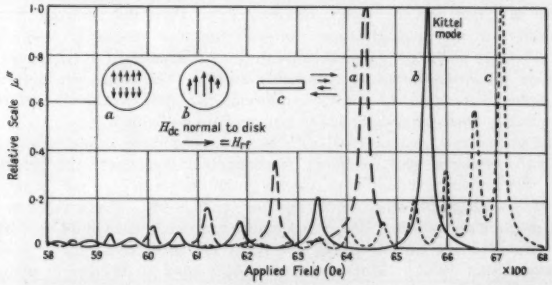


Fig. 16. Three series of non-uniform resonance modes observed at 9 Gc/s, with radio-frequency fields of different symmetries in a disk of Mn ferrite. Note that the Kittel, or uniform, mode is a member of only one series (*after* Dillon 1956, quoted by Walker 1958).

field at different relative phases. The energy of such a mode is not very different from that in which **M** has the same direction everywhere, as the variation with position is very slow (on an atomic scale), and there is therefore practically no change in the exchange energy. The only increase in energy which occurs is due to changes in internal demagnetizing fields, which arise from the long-range magnetic dipole forces of the spins. The magnitude of these internal fields is of the order of $4\pi M_s$, in general a few hundred gauss, and the field which must be applied to observe such a mode is therefore shifted from the uniform resonance by a readily accessible amount. Figure 16 shows some typical series of modes excited by radio-frequency driving fields with various symmetries. It is clear that this kind of resonance is fundamentally quite different from those involving coupled sub-lattices, and the reason why it has so far only been observed in ferrimagnetic crystals is that it is a bulk effect, which occurs only in spontaneously magnetic substances which are at the same time insulators.

The formal mathematical theory of these resonances was worked out by Walker (1957, 1958), who showed that there can exist several different series of modes, characterized by three integral mode numbers (n, m, r) which define the relative phases of the precession of **M** at different parts of the sample at all times. As the relative phases are constant in time the energies of the modes can be calculated in static terms, and the resonances are usually referred to as ' magnetostatic modes '.

The general relationship between the frequency and the applied field is rather complicated, and useful nomograms have been prepared by Walker (1957, 1958) and by Fletcher and Bell (1959), who also tabulate the energies and spatial distribution functions of all modes up to $n = 5$.

Comparison between the theory and experiment in general showed excellent agreement, but in a few cases small discrepancies ($\sim 2\%$) remained. That the position was not quite as simple as appeared at first was underlined by further experiments (White and Solt 1957, and Solt, White and Mercerau 1958), which showed that extra resonances could be observed even if the radio-frequency field was practically uniform throughout the sample, and that these could not be indexed according to Walker's description. Dillon (1958 d) found empirically that it was possible to suppress these new modes by ensuring that the radio-frequency field was truly uniform (by removing dielectric inhomogeneities caused by the sample mounting, see § 5.3) and also ensuring that accurately spherical, smooth samples were used. The significance of these factors was explained by Fletcher, Solt and Bell (1959) who pointed out that in the relatively ' large ' samples used in these experiments (to ensure an adequate gradient of radio-frequency field for the observation of Walker modes) electromagnetic propagation effects are not quite negligible (cf. § 5.3(*d*)), and the magnetostatic approximation is then not entirely valid. As a result of this the different modes are not quite independent of one another, and a given radio-frequency field may excite otherwise forbidden modes. Moreover if two of the modes happen to be almost degenerate in energy, any coupling between them may be enhanced, and the field for resonance of each of them may be shifted by an appreciable amount. These effects depend on the size of the sample, as observed, and they may be investigated by making measurements at different temperatures so as to vary the bulk magnetization. In this way Fletcher *et al.* were able to reconcile the experimental and theoretical resonance fields for 31 observed modes to an accuracy of better than 0·1%, and it appears that these resonances, which should properly be called ' pseudo-magnetostatic modes ', are now very well understood. An excellent review of the theory of these modes, and their applications, has been given by White (1960).

A different kind of non-uniform resonance has been predicted by Kittel (1958). He showed that if the anisotropy acting on the ions near the surface of a small (or thin) sample is appreciably different from that acting inside the sample, then it might be possible to excite, by means of a *uniform* radio-frequency field, *non-uniform* modes with a relatively rapid variation of **M** with position. The energy of such a mode would involve the exchange energy, and it would correspond to the excitation of long wavelength spin waves. So far there is no definite evidence for such resonances in any ferrimagnetic material, although it has been suggested that some anomalous microwave spectra observed by Jarrett and Waring (1958) in very small crystals of $NiMnO_3$ may have been due to this effect. For thin ferromagnetic alloy films such resonances have been observed by Seavey and Tannenwald (1958).

5.7. *Line Widths*

In all the dynamic processes which we have considered so far we have neglected (with one exception) any effects of damping, since these are usually small. However it is clear that in all cases there must exist some mechanism which dissipates

energy from the precessional motion of the spins, since it would otherwise be impossible ever to magnetize a substance by the application of a steady field. In any resonance experiment the effect of damping is to broaden the absorption line, and this can be measured in terms of the difference in fields between the points at which $\mu_e'' = \frac{1}{2}(\mu_e'')_{max}$ (see Fig. 11 ; note : occasionally ΔH is defined as half this amount). The reduction of line widths to a minimum is of considerable practical importance, not only to improve accuracy and resolution in resonance experiments, but also to make feasible new types of experiments and applications at high microwave powers (see next section).

A considerable amount of recent work on line widths has shown that they may depend on many different factors, some of which might be regarded as spurious, inasmuch as they would not occur in an ' ideal ' ferrimagnet. For example there may be considerable damping from eddy currents if there is any appreciable electrical conductivity due to mixed valencies (see § 2.1). Apart from a few exceptions such as Fe_3O_4 (see Galt, Yager and Merritt 1954) this effect may be eliminated by proper control of stoichiometry (Yager, Galt and Merritt 1955), though this is often difficult in practice. Another ' spurious ' contribution to line widths, the importance of which has been recognized only very recently, arises from the presence of certain chemical impurities. It is well known that some paramagnetic ions (e.g. Co^{2+}, Fe^{2+}, various rare earths) are very strongly coupled to a crystal lattice via unquenched orbital angular momentum, so that their relaxation time τ is very short at ordinary temperatures. If such an ion is present as an impurity it will be coupled by the exchange interaction to the relatively undamped bulk magnetization and so provide a rapid means of energy dissipation to the crystal lattice. The effect of this coupling on the g-factor was discussed in § 5.3(c). Its contribution to the line width has been estimated by Kittel (1959) and de Gennes, Kittel and Portis (1959), who showed that subject to certain approximations

$$\frac{\Delta H}{H_{res}} = \frac{N_B}{N_A}\frac{\omega_{AB}{}^2\tau}{kT}, \qquad \ldots\ldots(5.19)$$

where $\omega_{AB} = n(\gamma_A \mathbf{M}_B + \gamma_B \mathbf{M}_A)$ is the exchange frequency between the undamped magnetic moment M_A and the damping impurity M_B (cf. Eqn (5.10)), and N_A and N_B are the relative number of ions present. As $\hbar\omega_{AB} \gg kT$, while τ is generally not so short that $\omega_{AB}\tau \ll 1$, an appreciable contribution to ΔH can result from a relatively small number of impurity atoms. Thus Dillon and Nielsen (1959) found that 0·1% Tb^{3+} in yttrium iron garnet increases ΔH from a fraction of an oersted to several hundred oersted (see Fig. 17). Similarly, White (1959 b) pointed out that various anomalous line widths previously observed in ferrites (see for example Tannenwald 1956) may be explained by small impurities of Fe^{2+} known to have been present. The considerable effect of small Fe^{2+} impurities on ΔH has recently been demonstrated by Schnitzler, Folen and Rado (1960), who made measurements on a sample of $Li_{0.55}Fe_{2.45}O$ before and after subjecting it to oxidizing heat treatment and found a reduction by a factor of 2.

Equation (5.19) also explains the increase in ΔH frequently observed as the temperature is lowered, since for most paramagnetic ions τ increases rapidly at low temperatures. If τ increases so much that $\omega_{AB}\tau \sim 1$ an expression more complete

than Eqn (5.19) must be used (Kittel 1959) and a *decrease* of ΔH is then predicted as τ increases further, in agreement with experiment (see for example Fig. 17).

While it seems certain that these effects have masked the observation of true line widths in very many cases in the past, it is of interest to consider also some of the more 'intrinsic' mechanisms which have been identified as being important.

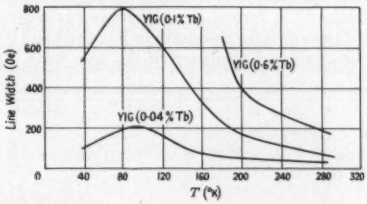


Fig. 17. Variation of line width with temperature for three samples of YIG containing different amounts of terbium impurity. The line width of pure YIG in this region is less than 0·4 Oe (*after* Dillon and Nielsen 1959).

The relaxation processes in any resonance experiment can generally be of two kinds: the dissipation of energy from the uniform precession can proceed either by means of a direct transfer of energy from the spins to the crystal lattice, or else by the excitation of other kinds of spin states which ultimately give up energy to the lattice. In strongly coupled systems, such as ferrimagnets, the spin–spin processes generally dominate over direct spin–lattice relaxation (Akhieser 1946, Kittel and Abrahams 1953) and most theories have attempted to describe how the resonance mode decays into more disordered spin states. It may be emphasized that these theories have therefore neglected *a priori* the possibility that certain kinds of ions may be coupled to the crystal lattice even more strongly than to the other spins, as discussed above, and so apply only to certain pure compounds. However this is in fact not a very serious restriction, as very many ferrimagnetic compounds contain only ions which are in S-states for which direct spin–lattice relaxation is very slow.

One method of finding ΔH in terms of the spin–spin coupling is the calculation of mean nth moments of the line shape about its centre (Van Vleck 1950, Keffer 1952). This is a good method above the Néel point, but it is poor at low temperatures where the collective state of the spins must be described in terms of spin waves. Early attempts to calculate the scattering of spin waves were faced with the difficulty that in the usual theory, as developed for the specific heat and magnetization, the uniform precession which corresponds to a wave vector $\mathbf{k} = 0$ has a smaller energy than any mode with $\mathbf{k} > 0$. Thus in order to excite a non-zero $\mathbf{k}$ spin wave energy would have to be provided from some source other than the uniform precession, and this would make such a process very unlikely (Kittel and Abrahams 1953, Kasuya 1954). This difficulty was resolved by Clogston *et al.* (1956), who pointed out that in a body of finite size it is necessary to take demagnetizing fields into account, and for a wide range of low $\mathbf{k}$ values these dominate the effects

of exchange on the spin wave spectrum. Considering specifically an ellipsoid of revolution in a magnetic field H_0 parallel to its z axis, Suhl (1957 a) had shown that the dispersion relation between angular frequency ω and wave vector $\mathbf{k}$ is given by

$$\omega_k^2 = \gamma^2(H_0 - N_z M_s + H_e a^2 k^2)(H_0 - N_z M_s + H_e a^2 k^2 + 4\pi M_s \sin^2\theta) \quad \ldots\ldots (5.20)$$

where N_z is the demagnetizing factor, H_e an effective exchange field, a the lattice constant of the crystal (assumed for simplicity to be cubic) and θ the angle between the z axis and the direction of propagation of the spin wave. (N.B.—$H_e a^2$ here

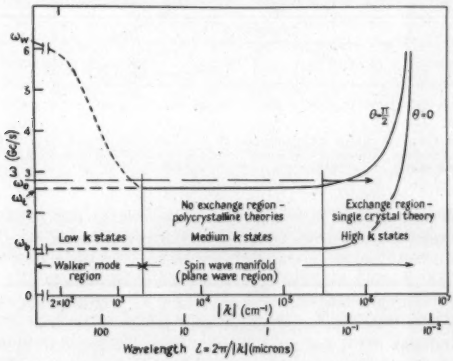


Fig. 18. The spin wave spectrum (Eqn (5.20)) for YIG, drawn approximately to scale assuming $4\pi M_s = 1800$ gauss, $H_0 = 1000$ Oe, $\gamma = 2{\cdot}8$ Mc/s Oe^{-1} ($g_0 = 2$); $H_e a^2 = 10^{-10}$ Oe cm^2, $N_z = N_t = 4\pi/3$ (sphere).

$$\omega_t = \gamma\{(H_0 - N_z M_s)(H_0 - N_z M_s + 4\pi M_s)\}$$
$$\omega_b = \gamma(H_0 - N_z M_s)$$
$$\omega_0 = \gamma\{H_0 - (N_z - N_t) M_s\}$$
$$\omega_W = \gamma(H_0 - N_z M_s + 2\pi M_s)$$

ω_t and ω_b are respectively the top and bottom of the spin wave spectrum extrapolated to $\mathbf{k} = 0$; ω_0 is the frequency of the uniform (Kittel) mode, here shown above ω_t. For higher fields ω_0 generally lies between ω_t and ω_b. ω_W is the upper limit of the Walker mode spectrum, which are the proper modes for $|\mathbf{k}|$ less than about 2000 cm^{-1} in a typical sample $\sim 400\mu$ in diameter (*after* Buffler 1959).

corresponds to $D/\hbar$ in Eqn (4.1).) An extension to the case of a general ellipsoid has been given by Callen (1958). The form of this relation shows that spin waves of a given $\mathbf{k}$ have energies extending over a band of frequencies, so that the uniform precession, whose frequency is given by the Kittel relation

$$\omega_0 = \gamma\left[H_0 + \frac{3M_s}{2}\left(\frac{4\pi}{3} - N_z\right)\right],$$

is now degenerate with a wide range of spin waves (see Fig. 18).

There are several mechanisms which can now couple the degenerate spin waves. The one considered by Clogston *et al.* was the effect of anisotropic exchange between ions in a disordered crystal structure, such as occurs in inverse spinels. Anisotropic exchange is large for ions having partially unquenched orbital angular momentum (see § 3.4), and an order of magnitude estimate showed that observed line widths of about 50 gauss could be accounted for in these cases (see also Callen and Pittelli 1960). However ions of this kind may also give rise to the direct relaxation process discussed above, so that they constitute a difficult case to treat quantitatively. For substances containing exclusively ions in S-states the only anisotropic coupling comes from magnetic dipole–dipole interactions, but these generally contribute

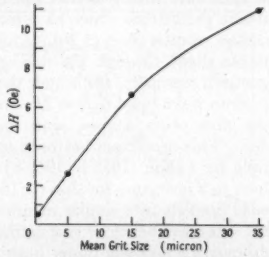


Fig. 19. Effect on line width of reducing surface roughness by using progressively finer polishing paper, for a single crystal sphere of YIG at 9300 Mc/s (*after* LeCraw, Spencer and Porter 1958 b).

only a small amount to the line width (White 1959 c). Sparks and Kittel (1960) have shown that dipole interactions are sufficient to account for the very narrow line widths ($\sim 0{\cdot}2$ Oe), which have been observed in extremely pure YIG at low temperatures (LeCraw and Spencer 1960), but generally they will be masked by other effects. Apart from the effects mentioned above, coupling may arise from spatial fluctuations in the crystalline electric field, or possibly from random variations of the exchange forces due to lattice imperfections or cation disorder. The effect of electromagnetic propagation has been considered by Gintsburg (1959), but his conclusions have been challenged by Clogston, Suhl and Walker (1959) who claim to show that this effect will be small for the size of samples usually measured.

Scattering of spin waves can also arise from macroscopic irregularities such as roughness of the sample surface (LeCraw, Spencer and Porter 1958 b, Spencer, LeCraw and Clogston 1959). This process is somewhat different from those involved in atomic scattering, since the coupling is then to modes of relatively low **k** (long wavelength), comparable with the scale of the irregularities. The striking effect of removing surface irregularities by progressive polishing is demonstrated in Fig. 19. On the other hand really gross deformations of the sample surface affect the line width only slightly (Geschwind and Clogston 1957) as there are relatively few modes with ' wavelengths ' long enough to be affected. Thus the non-uniform internal fields produced by actually truncating a spherical sample may cause a few discrete magnetostatic modes to be excited, but the uniform resonance mode is

practically unaffected. In the same way the line width caused by polycrystalline structure involves only modes of low **k** and is therefore smaller than would be predicted from a straight average of the anisotropy fields (see § 5.4).

One of the difficulties in all these scattering calculations is that all spin wave modes do not appear to be equally effective in dissipating energy, in that some are more tightly coupled than others to the acoustic modes of the lattice. This was demonstrated qualitatively by White (1959 a), who investigated the variation of line width with mode number for various magnetostatic modes. So far there is no way of calculating the preferential weight which should be given to any particular group of degenerate modes, and it is therefore difficult to make any exact quantitative calculations. However the spin wave theory as developed by Clogston *et al.* does make several definite qualitative predictions which have found striking verification. From the form of the dispersion relation (Eqn (5.20)) it follows that ΔH should be strongly dependent on sample shape through the demagnetizing factor N_z. In particular a thin disk magnetized normally, for which the Kittel mode coincides with the very bottom of the spin wave band $\omega_b/\gamma = H_0 - 4\pi M_s$, should have a line width appreciably narrower than other shapes, since in this case there is no degeneracy with other modes. This prediction has been confirmed experimentally in several cases, for example by Dillon (1957 b, 1960 b) who found changes in ΔH from 25 gauss for a sphere to a few gauss for thin (1 : 16) disks.

The spin wave theory also predicts in a similar manner that there should be a discontinuous change in ΔH if the frequency or magnetization is varied in such a way that the uniform mode moves above the upper limit of the $\mathbf{k} = 0$ spin wave spectrum (see Fig. 18). For a sphere the condition for this is $\omega = \gamma H_0 < 8\pi\gamma M_s/3$. This has been verified experimentally by Buffler (1959, 1960). Further confirmation of the general ideas of a finite spin wave manifold is found in connection with high power effects (see § 5.8).

Relaxation processes are often allowed for empirically by adding damping terms to the classical equations of motion of the magnetization vector. These take the form of torques **R** which may be expressed either in the form

$$\mathbf{R} = -\frac{\lambda \mathbf{M} \times (\mathbf{M} \times \mathbf{H})}{M^2} \qquad \ldots\ldots(5.21)$$

(Landau and Lifshitz 1935) in which the magnitude but not the direction of **M** stays constant, or in the form

$$R_x = -M_x/\gamma_e T_2\,; \quad R_y = -M_y/\gamma_e T_2\,; \quad R_z = -(M_z - M_0)/\gamma_e T_1 \qquad \ldots\ldots(5.22)$$

where T_1 and T_2 are the ' longitudinal ' and ' transverse ' relaxation times (Bloch 1946, Bloembergen 1950). A general discussion and extension of these forms, and their relation to the general spin wave theory, has been given by Callen (1958).

In most treatments of damping, essentially *ferri*magnetic multi-sub-lattice effects have so far been ignored, and the theory has been developed as for ferromagnetic resonance in insulators. Wangsness (1958) considered the effect of introducing macroscopic damping terms for each individual sub-lattice and found that physically reasonable results could be obtained only by using terms of the form of Eqn (5.21). For small damping the combined effect of the various sub-lattices

can then be expressed in terms of a suitable average damping term acting on the resultant magnetization. If, on the other hand, one of the sub-lattices is highly damped completely new solutions are found for the equations of motion. An extreme limit of this case has already been discussed above in connection with substances containing certain impurities (Eqn (5.19)).

5.8. *High Power Effects*

Further information about relaxation processes can be obtained from resonance experiments at power levels which are comparable with the maximum rate at which energy can be dissipated by the magnetic system. As the radio-frequency power

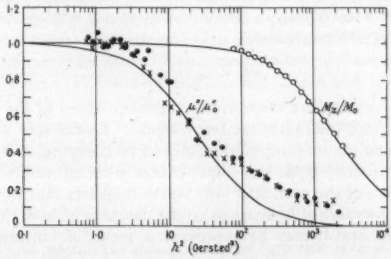


Fig. 20. Saturation of μ'' and M_z as a function of radio-frequency power, at a fixed value of static field, for a sample of Ni ferrite ($h \propto$(r.f. power)$^{\frac{1}{2}}$). Two successive runs for μ'' are shown. The ordinates are normalized to the values at $h = 0$ (*after* Damon 1953).

increases we may expect to find two effects : (i) the height of the peak in μ'' (which is proportional to the fraction of power absorbed) will decrease, and (ii) as the uniform precession increases the average component of magnetization M_z in the direction of the applied d.c. field will decrease. From the classical equations of motion one can predict that both these effects should set in at the same power level, given by the condition

$$\tfrac{1}{4}\gamma_e^2 h_c^2 T_1 T_2 \sim 1$$

where h_c is the corresponding radio-frequency field amplitude and T_1 and T_2 are defined by Eqn (5.22) (Bloch 1946). If, as is often assumed,

$$T_1 \simeq T_2 \simeq 2/\gamma_e \Delta H \qquad \ldots\ldots(5.23)$$

this condition becomes $h_c \sim \Delta H$.

Measurements of μ'' and M_z as a function of h have been carried out on a number of ferrimagnetic substances by Damon (1953) and Bloembergen and Wang (1954), who found striking discrepancies with this prediction in all cases. They observed that while M_z behaves very much as expected, saturation of μ'' sets in at powers about 100 times lower (see Fig. 20), in contrast to the behaviour of paramagnetic specimens for which μ'' and M_z decreased together. They found, moreover, that as h was increased an unexpected secondary peak in μ'' appeared at an applied field somewhat below that of the main resonance.

Explanations of these effects have been given by Suhl (1956 a, b, 1957 a, 1958, 1959); for a brief review of the theory see Suhl (1958). He showed that both the premature decrease in μ'' and also the secondary peak are due to the excitation of certain spin wave modes which are coupled to the uniform precession. For both effects the coupling is produced mainly by long-range magnetic interactions, rather than exchange forces, and the spin waves excited are those with small $\mathbf{k}$, i.e. long wavelength. Indeed in some cases the wavelengths may become so long that the modes should properly be described in terms of the magnetostatic modes (see § 5.5), but so far this complication has only been considered qualitatively (see Suhl 1958). The particular spin waves excited in the two effects are different.

In a uniform crystal the saturation of μ'' arises from pairs of spin waves having equal and opposite $\mathbf{k}$ travelling in the direction of the applied field, and subject to the energy conservation condition

$$\omega_{\mathbf{k}} = \omega_0 = \omega_s,$$

where $\omega_{\mathbf{k}} \equiv \omega_{-\mathbf{k}}$, ω_0 and ω_s are respectively the frequencies of the spin waves, the uniform precession and the radio-frequency signal. As the spin waves are excited their back reaction on the uniform mode increases its damping, and so decreases μ''. For small radio-frequency fields this effect is negligible but above a certain critical field h_c the excitation of the particular spin waves increases rapidly and a change in μ'' is observed. The value of h_c depends on the damping of both the uniform mode and the spin waves, and it may be expressed in terms of 'intrinsic line widths' ΔH_0 and ΔH_k as

$$h_c \simeq \Delta H_0 \left(\frac{\Delta H_k}{4\pi M_s}\right)^{1/2}. \qquad \ldots\ldots(5.24)$$

ΔH_0 may be identified qualitatively with the observed line width of the uniform mode, while ΔH_k can only be inferred from measurements of h_c. Both ΔH_0 and ΔH_k may be affected by surface roughness or impurities but it is reasonable to assume that both will be of the order of, or less than, the line width measured at low powers. This is generally very much smaller than $4\pi M_s$ and Eqn (5.24) therefore shows why saturation of μ'' sets in at signal strengths very much less than ΔH, as observed by Damon (1953) and Bloembergen and Wang (1954). The average magnetization is practically unaffected by the spin waves, and so shows no saturation up to much higher powers.

Quantitative experimental evidence for the validity of Eqn (5.24) has been found by LeCraw, Spencer and Porter (1958 a) who measured h_c in single crystals of YIG, for which direct measurements of the relaxation time T_1 had been made (Farrar 1958). To avoid possible surface effects they used the value of $2/\gamma T_1$ instead of the measured line width for both ΔH_0 and ΔH_k, and found agreement with the experimental value of h_c to about 10%. The advantage of this procedure was emphasized by subsequent experiments (LeCraw and Spencer 1959) which showed that ΔH_k is indeed independent of the surface state of the sample, while the measured line width may easily be increased by a factor of ten. The complicated effects of surface roughness have been further investigated by Fletcher, LeCraw and Spencer (1960). They made measurements of T_1 and T_2 (using a novel frequency modulation method) as well as the line width and h_c, all on the same sample of YIG, under

various controlled conditions of surface roughness. From these data they were able to separate the effects of various relaxation processes, and in particular they showed how it was possible to obtain information on the intrinsic spin–lattice coupling of the system, even when surface effects are large. The effect of direct spin–lattice coupling by rare earth ions in garnets has been studied by Schlömann, Green and Milano (1960).

The reason why spin waves are created in pairs with equal and opposite **k** is that in a perfect crystal momentum must be conserved in any spin wave scattering process. However, Schlömann (1959 b) pointed out that this is not necessarily so in an irregular or imperfect crystal, in which momentum can be taken up by the lattice. It is then possible for the uniform precession to generate single spin waves provided that energy is still conserved. Such a process has no definite threshold, and μ'' will then decrease continuously with power (Suhl 1959), though in practice the initial variation may be slow. Experiments under these conditions have been carried out by Weiss (1960) and by Schlömann, Green and Milano (1960). The latter found that if the irregularities were on a very small scale (corresponding to the excitation of very high **k** spin waves) it is even possible for μ'' to *increase* as a function of h, before finally falling off as $1/h^2$.

The appearance of the subsidiary resonance is explained by the growth of pairs of spin waves, **k** and $-\mathbf{k}$, having energies given by

$$\omega_s = 2\omega_k. \qquad \ldots\ldots(5.25)$$

In this process one microwave quantum splits into two spin waves, in contrast to the first process in which two quanta give rise to two spin waves, and the inelastic scattering process in which one quantum produces one spin wave. The range of possible values of ω_k is given by the dispersion relation (Eqn (5.20)) which shows that this process can be observed only for applied fields H_0 subject to

$$H_0 < \tfrac{1}{2}\omega_s/\gamma_e + 4\pi M_s N_z. \qquad \ldots\ldots(5.26)$$

In general this requires $H_0 < H_{res}$, and a detailed calculation shows that the potentially most unstable spin waves will generally be created when H_0 is a few hundred gauss below the uniform resonance field. These will be spin waves travelling at some oblique direction to H_0 and their effect will be to increase the losses at this field and so give rise to a peak in μ''.

A particularly interesting situation arises when the experimental conditions are such that

$$4\pi M_s N_z \geqslant \tfrac{1}{2}\omega_s/\gamma_e, \qquad \ldots\ldots(5.27)$$

since the resonance conditions for the main and the subsidiary peak can then be satisfied simultaneously, i.e.

$$\omega_s = \omega_0 = 2\omega_k. \qquad \ldots\ldots(5.28)$$

The critical field for the onset of instabilities is then given by

$$h_c \simeq \frac{\Delta H_0 \Delta H_k}{4\pi M_s}, \qquad \ldots\ldots(5.29)$$

which for materials with narrow line widths is very considerably less than the non-coincidence threshold (Eqn (5.24)). This situation was investigated experimentally

in some detail by Spencer and LeCraw (1958), Spencer, LeCraw and Porter (1958) and LeCraw, Spencer and Porter (1958 a), who measured h_c for single crystal spheres of YIG ($\Delta H_0 = 2$ Oe, $4\pi M_s \simeq 1725$ gauss), both as a function of frequency and as a function of M_s (by varying the temperature). In both cases they found extremely low threshold fields (~ 2 to 4 mOe as against 46 mOe) when the coincidence condition (Eqn (5.29)) was satisfied. For purer samples with narrower line widths the reduction should be even bigger (LeCraw and Spencer 1960).

The response to radio-frequency fields larger than h_c is very complicated, especially for the case of main resonance line. Schlömann and Green (1959) showed that as the radio-frequency power is increased μ'' does not decrease smoothly, but rather in a series of small discontinuous steps. These have been interpreted by Schlömann (1960 a) as due to exact coincidences in frequency and direction of the potentially unstable spin waves with the driving frequency and the d.c. field respectively. At high powers the spin wave frequencies vary slightly due to changes in the demagnetizing fields, and exact coincidences then become possible. The behaviour of the secondary resonance at high powers has been discussed by Suhl (1956 a, 1957 a), who showed that a rather exact description was possible for this case. At powers well above the first threshold further complications set in (Weiss 1960, Suhl 1960).

Several other new effects have been observed at high power levels. Spencer and LeCraw (1958, 1959) found that magneto-acoustic resonances could be induced in small samples of YIG by microwaves at either 9000 or 3000 Mc/s. For the size of samples used the acoustic resonances lay between 5 and 30 Mc/s, and they could be induced with the minimum radio-frequency power by applying a d.c. magnetic field such that the uniform precession mode differed from the signal frequency by just the acoustic frequency. A superficially similar effect was found by Weiss (1958, 1959) who observed radio-frequency oscillation of the magnetization at frequencies between 0·1 and 15 Mc/s under conditions of non-linear microwave excitation of the uniform mode. These oscillations could also combine with the signal frequency to produce a series of side-bands centred about the main resonance. The effect has been explained in terms of a ' fold-over ' in the variation of precession angle with magnetic field (Anderson and Suhl 1955, Suhl 1956 a, b, 1960, Weiss 1958, 1959). This again arises from the critical excitation of certain spin waves, which in this case cause an apparent dependence of magnetization or anisotropy on radio-frequency power (see also Masters 1960, Gottlieb 1960). Another interesting non-linear effect has recently been observed by Schlömann, Green and Milano (1960), in experiments in which the d.c. and radio-frequency fields were aligned parallel (see also § 8.3).

Apart from their fundamental interest, the importance of all these high power effects lies in their potential application to microwave parametric amplifiers and oscillators (Suhl 1957 b, c, 1958, Weiss 1957). In this class of devices, power is supplied at one frequency (usually the uniform precession mode) and amplification is achieved of some other frequency chosen to correspond to a mode coupled non-linearly to the driven precession. There are various ways of operating such devices and a good review of several of these can be found in Suhl (1958).

5.9. *Dynamic Effects in Unsaturated Samples*

In this section we shall consider briefly three classes of phenomena which are observed when dynamic measurements are made in the presence of applied fields too weak to remove domain structure.

5.9.1. *Multiple microwave resonances.*

The effect of domain structure on microwave resonance was first discussed by Polder and Smit (1953) and by Nagamiya (1953). The general situation is very complicated and explicit calculations can only be made for certain simple domain configurations. The magnetization within each domain precesses under the combined effect of the applied anisotropy and demagnetizing fields, and in general the various domains are not in phase. This gives rise to dynamic demagnetizing effects at the Bloch walls (as well as on the surface of the sample), and these depend both on the domain shapes and the direction of the exciting radio-frequency field. As a result of the demagnetizing effects the resonance splits into two or more separate lines. Effects of this kind have been observed in cobalt ferrite (Tannenwald 1955 a), in manganese ferrite at low temperatures (Tannenwald 1955 b), in magnetite below its transition (Bickford 1950), in rare earth garnets at low temperatures (Jones, Rodrigue and Wolf 1958) and in uniaxial $BaFe_{12}O_{19}$ (Smit and Beljers 1955). A general discussion of the theory has been given by Smit and Beljers (1955) and by Artman (1956, 1957 a). Qualitatively one may expect multiple resonances from this cause whenever the anisotropy becomes comparable with the applied d.c. field. Conversely, for a given anisotropy, it should always be possible to suppress them by observing the resonance at a higher frequency and field, or else applying the field along one of the 'easy' axes of magnetization.

5.9.2. *Domain wall motion.*

On account of their low electrical losses ferrimagnetic substances are well suited for the study of the dynamic changes which take place when a domain structure is altered by the application of a magnetic field. In single crystals the domain walls are relatively free, and they move with a velocity which depends, among other factors, on the damping of the spin system, and which should therefore be related to the measured resonance line width. In fact the two may be expected to be inversely proportional, unless there exist damping mechanisms whose characteristic time is long compared with microwave frequencies, but short compared with the time of domain motion. Such mechanisms would affect the velocity but not ΔH. They can arise from various causes including electron interchanges (i.e. electrical conduction) between ions in different valency states (e.g. $Fe^{2+} \rightleftarrows Fe^{3+}$), and rapidly relaxing magnetic ions (see § 5.7). Measurements of the domain wall velocity can therefore provide important additional information on the fundamental relaxation processes. Unfortunately there are considerable experimental difficulties and only a few such measurements have been reported so far (Galt 1952, 1954, Galt, Andrus and Hopper 1953, Dillon and Earl 1959 a). A review of the methods and theory has been given by Galt (1957). In general the velocities found at high temperatures agree with those calculated from the line widths, but at very low temperatures there are big discrepancies (~ 40 times). This is consistent with the

fact that line widths often pass through a maximum which is ascribed to an increase in some ionic relaxation time at low temperatures (see § 5.7).

5.9.3. *Initial permeability.*

A technically very important property of magnetic materials is the response to small alternating fields in the absence of any d.c. field, characterized by the initial permeability μ_{init}. In general μ_{init} is complex and its components depend strongly on frequency. Many measurements of μ_{init} have been reported in the literature, and there has been much discussion and controversy about the magnetization processes for various materials and frequencies. Brief general reviews of the position have been given by Bagguley and Owen (1957) and by Birks (1957).

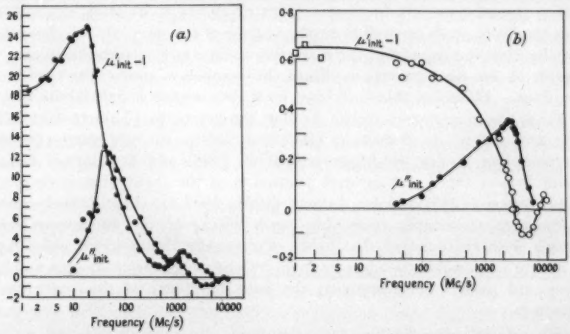


Fig. 21. Initial permeability plotted against frequency for a commercial ferrite (ferramic A): (*a*) sintered sample, showing two dispersion regions, (*b*) same substance powdered and dispersed in wax, showing only high frequency dispersion (*after* Rado, Wright and Emerson 1950). Note the change of scales.

In general one may expect two mechanisms to contribute to the initial permeability: (i) motion of the domain walls, and (ii) rotation of the magnetization within the domains against the anisotropy. It seems to be agreed that the second of these mechanisms is to some extent important in all cases, as suggested originally by Snoek (1947 b, 1948), and the main question has been to decide under what conditions wall motion is also important. Rado and his co-workers (Rado, Wright and Emerson 1950, Rado and Terris 1951, Rado, Wright, Emerson and Terris 1952, Rado 1953, Rado, Folen and Emerson 1957), in a series of experiments on various magnesium ferrites, showed that rotation against anisotropy was certainly not sufficient in these cases to account for the observed values of μ_{init}' at relatively low frequencies, and they also found two regions of frequency in which the losses μ_{init}'' were large. The one at higher frequencies they could identify quantitatively as due to rotations, and the one at low frequencies they ascribed to wall motion. In support of this conclusion they carried out experiments in which the same sample was measured in sintered and finely powdered states—the aim being to influence the effect of wall motion—and they found that the low frequency relaxation was indeed suppressed by powdering (cf. Figs 21 (*a*) and (*b*)).

On the other hand several other groups, notably that at the Philips Laboratory, Eindhoven, working mainly on various nickel ferrites (Snoek 1948, Went and Wijn 1951, Wijn, Gevers and van der Burgt 1953, see also Smit and Wijn 1954), found only one dispersion region, and concluded that rotation was primarily responsible for initial permeability in these cases. They argued that reversible domain wall motion is not possible even in sintered polycrystalline materials such as ferrites, and they showed that the frequency of the dispersion region was consistent with that calculated from Snoek's theory, assuming rotation only, and the measured value of μ_{init}' at zero frequency.

A recent series of experiments to reconcile these divergent views has been reported by Pippin and Hogan (1959, see also Pippin 1959), who also review the earlier work. Investigating several different series of ferrites and garnets with well-controlled properties they showed that it was possible to observe two dispersion regions for most materials, but only if the density was sufficiently high ($\sim$90 to 95% of the single crystal density). From the absolute magnitudes of μ_{init}' they confirmed the conclusion that in these cases rotation alone is not able to account for the magnetization in the radio-frequency range. For somewhat lower densities however, the two dispersion regions coalesce to one, presumably because domain wall motion is impeded by air pores and grain boundaries (see also Kriessman, Harrison and Black 1958).

At very high frequencies ($\sim$100 to 1000 Mc/s) magnetization by rotation is the only possible mechanism, and this controls the maximum frequency to which a material can be used with an appreciable permeability. Polder and Smit (1953) showed that absorption (and dispersion) by rotation against the anisotropy extend over a range of frequencies $\gamma_0 H^a < \omega < \gamma_0(H^a + 4\pi M_s)$ because of dynamic demagnetizing effects on the domain walls, and in most cases the upper limit of this gives the maximum usable frequency. For this frequency to be high the anisotropy must be large, but this implies that the magnitude of the rotational permeability is correspondingly low. An important recent development to overcome this difficulty has been the discovery of the new hexagonal oxides (see § 2.2(d)) some of which have a large H^a in one direction, but low H^a in a perpendicular plane (see § 3.4). Rotational permeability which is inversely proportional to the small H^a is then large, but the dispersion region is still quite high as the frequency in this case is of order $\gamma_0(H_\parallel{}^a H_\perp{}^a)^{1/2}$, and $H_\parallel{}^a$ may be very large. Such materials are of great technical importance: for further details see Smit and Wijn (1959).

§ 6. Experiments with Neutrons

6.1. *Neutron Diffraction*

When thermal neutrons pass through an ordered magnetic material there is an interaction between their magnetic moments and those of the material. This can be observed in several ways. If the neutrons are mono-energetic, and the sample polycrystalline, Debye–Scherrer type diffraction will occur, very much as for x-rays which have comparable wavelengths ($\sim$1Å). The diffracted intensity at various angles will be proportional to the sum of two factors, one depending on the nuclear scattering from the whole crystal lattice, the other on the magnetic scattering from the ionic moments. If all the ions in a particular crystallographic site are

magnetically similar, i.e. aligned parallel, magnetic scattering does not give rise to any extra lines, but only changes the intensities of lines already present from nuclear scattering. Thus the transition from paramagnetism to ferrimagnetism involving the ordering of ions already crystallographically inequivalent is not nearly as striking as the effects usually observed for antiferromagnets (see for example Shull and Smart 1949, Shull, Strauser and Wollan 1951). However by making quantitative intensity measurements some very direct evidence of ferrimagnetic ordering can be obtained.

There are basically three ways of distinguishing nuclear from magnetic scattering: (i) If the temperature of the sample is raised to above its Néel point coherent magnetic scattering disappears, leaving only the nuclear part. The effect of this is changed only slightly by temperature and in a way for which a correction can easily be applied. (ii) Magnetic scattering is also removed if the ionic magnetic moments are made parallel (or antiparallel) to the scattering vector. If anisotropy is not too large this can be achieved by applying a magnetic field. (iii) The contribution from magnetic scattering varies strongly with the scattering angle (according to a magnetic 'form factor') and it becomes very small for high order reflections. From the intensity of these it is then possible to calculate the nuclear contribution for the lower order reflections, whence the magnetic scattering can again be found by taking differences from the total intensities. For a general discussion of neutron diffraction see, for example, Bacon (1955) or Shull and Wollan (1956).

The magnetic scattering at various angles may be compared with that predicted by any model of ordered magnetic moments, and one of the first confirmations of Néel's theory was obtained in this way for Fe_3O_4 (Shull, Wollan and Koehler 1951). Many subsequent neutron experiments have confirmed the theory in other ferrimagnetic materials (e.g. Corliss, Hastings and Brockman 1953, $MgFe_2O_4$, Bertaut *et al.* 1956, YIG, Prince 1957 b, YIG, Goedkoop, Hvoslef and Živadinović 1959, $Ba_2Zn_2Fe_{12}O_{22}$). It must be pointed out however that the accuracy has not always been high enough to rule out slightly canted spin arrangements (see for example Hastings and Corliss 1953). Definite evidence for non-collinear spin arrangements has been found in a few cases (e.g. Prince 1957 a, $CuCr_2O_4$, Kasper 1959, Mn_3O_4), in agreement with magnetic measurements.

In some other cases neutron diffraction has apparently been at variance with the magnetic data (Pickart and Nathans 1958, 1959, $Ni(Fe,Cr)O_4$, $Mn(Fe,Cr)O_4$). One possible explanation for these discrepancies is that neutron diffraction is not sensitive to random deviations from a collinear ordered arrangement, while these can, on the other hand, contribute to high field susceptibilities and the magnitude of the saturation moment (see §§ 3.2 and 3.3). The exact cause for any such random deviations does not seem to be understood at present, though they could well be connected with local crystal field effects. Anisotropic local environments could also influence the neutron scattering by affecting the magnetic form factor, making it dependent on the scattering direction relative to the local symmetry axes. They would also be important indirectly if a magnetic field is used to separate the nuclear and magnetic scattering ((ii) above), since it may then not be possible for all the spins to become aligned parallel to the scattering vector. At present one can only conclude that, while neutron diffraction contributes important confirmatory evidence

for theories of magnetic ordering, any apparent discrepancies must be considered with great care. The use of single crystal samples in conjunction with polarized neutrons will undoubtedly throw light on some of these problems, but so far no such measurements seem to have been reported.

From the intensity of the magnetic scattering it is also possible to deduce average magnetizations of the individual sub-lattices and their variation with temperature. There are considerable experimental difficulties in obtaining very accurate values in this way, and the results which have been obtained so far can really only be described as 'consistent with' the values obtained magnetically ($\sim 10\%$) (e.g. Hastings and Corliss 1956). However one may hope that more precise neutron experiments will ultimately lead to considerably more reliable values for the sub-lattice magnetizations since these would be direct measurements, not dependent on any particular theory of interactions such as the molecular field approximation.

Neutron diffraction has also been used to obtain important crystallographic information in ferrimagnetic materials. Due to considerable differences in the magnetic scattering cross section for various ions it is possible to deduce ion preferences for inequivalent crystallographic sites, such as occur in the spinels (§ 2.1) (Bacon and Roberts 1953, $MgFe_2O_4$, $(Mg,Al)Fe_2O_4$, Corliss, Hastings and Brockmann 1953, $MgFe_2O_4$, Hastings and Corliss 1956, $MnFe_2O_4$, Nathans *et al.* 1957, $(Mn,Mg)Fe_2O_4$). It is also possible to determine with high precision the positions of light anions, such as O^{2-}, which scatter x-rays only weakly (Corliss, Hastings and Brockman 1953, Hastings and Corliss 1956, Prince 1957 b, Pickart and Nathans 1959). Such measurements are important for magnetic work to determine the strength and symmetry of the crystal field at the cation sites.

Other applications of neutron diffraction include the study of magnetic annealing in non-stoichiometric cobalt ferrite (Prince 1956) (see § 3.5), and the electronic ordering which occurs in Fe_3O_4 at 119°K (Hamilton 1958) (see § 2.2).

6.2. *Inelastic Neutron Scattering*

Another important class of experiments with thermal neutrons, which has only recently been applied to the study of ferrimagnetism, is the inelastic scattering of thermal neutrons by spin waves (Elliott and Lowde 1955, Brockhouse 1957, 1958, Riste, Blinowski and Janik 1959). If a beam of mono-energetic neutrons passes through an ordered magnetic material whose equilibrium state can be described in terms of a spin wave (or magnon) spectrum, an exchange of energy can take place between the neutrons and the magnons. This can be observed either directly by measuring the change in energy of the scattered neutrons, or else from the diffuse scattering at angles close to the Bragg diffraction peaks (see for example Elliott and Lowde 1955). By far the most likely scattering process corresponds to the excitation or annihilation of a single spin wave quantum by a single neutron, and for this momentum must be conserved according to the relation

$$\mathbf{k}_s - \mathbf{k}_i = \mathbf{k} + 2\pi\boldsymbol{\tau}$$

where $\mathbf{k}_i$ and $\mathbf{k}_s$ are the wave vectors of the incident and scattered neutron, $\mathbf{k}$ that of the spin wave and $\boldsymbol{\tau}$ is a reciprocal lattice vector with magnitude $1/d$, where d is

the relevant interplanar spacing in the crystal lattice. The magnitudes of $\mathbf{k}_i$ and $\mathbf{k}_s$ are related to the kinetic energy of the neutron

$$E_i = \frac{\hbar^2 k_i^2}{2m_0}, \quad E_s = \frac{\hbar^2 k_s^2}{2m_0},$$

where m_0 is the mass of the neutron. Conservation of the total energy demands that

$$E_i - E_s = \pm \hbar\omega_k$$

where $\hbar\omega_k$ is the energy of the spin wave involved. It is therefore possible to investigate directly the form of the dispersion relation between ω_k and k, and compare it with that calculated theoretically, or inferred indirectly from specific heat or magnetization data (§§ 3.2 and 4.1). With the range of neutron energies available at present it is not possible to investigate the low energy region where demagnetizing effects become important (Eqn (5.20)), and the first experiments aimed at verifying the validity of the k^2 law (Eqn (4.1)) and determining the exchange constant D (Brockhouse 1957, 1958).

Although this is a potentially very powerful method there are serious experimental difficulties, and only semi-quantitative results have been obtained up to the present. All the experiments have so far been carried out on one substance, Fe_3O_4, and for this the k^2 dispersion law has been verified. The value of the constant D found was about four times bigger than that deduced from the specific heat, but it was in more reasonable agreement with that estimated from the Néel temperature. One interesting feature demonstrated by the experiments of Riste *et al.* was that the spin wave theory is a good approximation up to surprisingly high temperatures ($\sim 0{\cdot}9\ T_N$), the exact limit depending on the energy of the particular magnons being excited. This is in agreement with the theory of Dyson (1956), which predicts that magnon–magnon interactions are relatively unimportant, particularly for long wavelength (i.e. low energy) spin waves. It may be expected that much further information will be obtained from neutron experiments if some of the experimental difficulties can be overcome.

§7. Optical Properties

7.1. *Absorption and Faraday Rotation*

On account of their low electrical losses at very high frequencies some ferrimagnetic materials are transparent even in the optical range. For all substances investigated so far the absorption is still somewhat greater than for real dielectrics, and thin ($\sim 0{\cdot}005$ cm) single crystal sections must be used for visual observation of the transmission. With such samples it is possible to carry out several simple optical experiments which are important both from a fundamental and a practical point of view. Dillon (1957 c, 1958 a, b, 1959) first observed that various rare earth garnets, notably YIG, have several rather broad transmission and absorption bands in the visible spectrum (see Fig. 22). These have been identified by Clogston (1959, 1960) as electric dipole transitions between electronic states of the Fe^{3+} ions, for which the normal selection rules may be relaxed by lattice vibrations. Garnets containing magnetic rare earth ions showed in addition some rather sharp absorption lines which may also be identified with electronic transitions. Important magnetic information can be obtained from such measurements. Thus, Wickersheim and

White (1960) have recently observed fine structure in the optical absorption of ytterbium iron garnet and interpreted this in terms of a splitting of the Yb^{3+} ground state by the exchange field of the Fe^{3+} sub-lattices. By applying an external magnetic field they were able to vary the direction of the exchange field and found a surprisingly large anisotropy in the splitting, indicating that the interaction in this case is more complicated than the simple scalar form of Eqn (2.1). Further optical experiments are reported to be in progress.

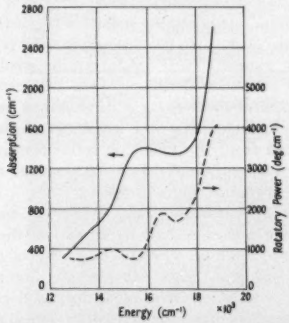


Fig. 22. Variation of the absorption and specific Faraday rotation for YIG in the visible range. Note that these two curves reflect the structure in each other (*after* Dillon 1958 a).

The optical absorptions are also of special interest because associated with each transition it is possible to observe Faraday rotation for plane polarized light propagated parallel to the direction of spontaneous magnetization. This is a non-reciprocal effect which is reversed if the direction of $\mathbf{M}_s$ is reversed, and this makes it possible to observe magnetic domain structures (see next section), while its magnitude gives further information on the electronic states of the magnetic ions (Clogston 1959, 1960). The variation of Faraday rotation with frequency for YIG is shown in Fig. 22. It can be seen that the effect is quite large, even for very thin samples. Associated with the Faraday rotation Dillon has also observed circular dichroism (preferential absorption ($\sim 10\%$) of light circularly polarized in one sense). For light propagated perpendicularly to the direction of $\mathbf{M}_s$, he observed magnetic birefringence (incident plane polarized light emerging elliptically polarized). Transmission and absorption measurements in the visible range have also been reported by Miles (1957, $MgFe_2O_4$ and SmIG), Kooy (1958, $BaFe_{12}O_{19}$) and Bailey (1960, various rare earth garnets).

Several infra-red absorption measurements have been reported (Waldron 1955, Wood 1959, various spinel ferrites, Porter, Spencer and LeCraw 1958, Wickersheim, Lefever and Hanking 1960, YIG). These absorptions depend mainly on crystal lattice vibrations, though Wickersheim *et al.* showed that in the case of the garnets impurities (e.g. Si^{4+}) are also important. An application of infra-red transmission has recently been reported by Buehler and Tanenbaum (1960), who used it to examine bulk crystals of YIG for flaws and inclusions of foreign phases.

7.2. *Observation of Domains*

An important practical application of the optical Faraday effect has been the direct observation of magnetic domains (Dillon 1958 a, b, 1959, Kooy 1958, Sherwood, Remeika and Williams 1959). If a suitably prepared single crystal section is examined under a polarizing microscope a pattern similar to Fig. 23 (*a*) or (*c*) (Plate I) is generally observed. The contrasting regions correspond to domains magnetized in different directions, the maximum contrast being obtained for simple 180° domains, aligned parallel and antiparallel to the line of sight. As the analyser in the microscope is rotated relative to the polarizer either the ' up ' or the ' down ' domains can be extinguished in turn (cf. Fig. 23 (*a*) and (*c*)) and for one position (polarizer and analyser parallel) both kinds can be made to appear the same (Fig. 23 (*b*)). In this last case the Bloch walls between the domains become visible as dark lines, since for these there is in general a net optical rotation away from the analysing plane. Domains whose magnetization is not parallel to the line of sight can be identified from the magnetic birefringence which they produce. Motion of the domains in an applied field can be observed (see Fig. 24 (Plate II)), and so can their disappearance at the Néel point.

The potential importance of these domain observations is that they are made on the whole sample, and not merely its surface as in the usual powder pattern techniques. However the samples which are used are of necessity very thin and their domain structures will still depend to some extent on the surface state, unless all strains are removed. This is not easy, as the normal electropolishing techniques cannot be used, and for garnets there are as yet no chemical etches, but it is possible to observe genuine bulk domains with careful mechanical polishing. Examination of the true domain structure gives information on the uniformity (chemical and physical) of the crystal samples, which is useful for other experiments such as microwave resonance.

Practical applications of these effects include measurements of switching times (important for computer cores), and field actuated light modulators, possibly even at microwave frequencies (Dillon 1958 a, b, 1959, Porter, Spencer and LeCraw 1958).

§ 8. Recent Developments

The main body of this report was completed in June 1960 and an attempt was made to include all the major references published up to that date. Since then there have been several important developments which we shall now review briefly. Together with the notes added in proof we hope that this will complete the report up to the end of 1960, though there may still be some important omissions, especially of work published in foreign languages.

8.1. *Complex Spin Configurations*

There now seems to be no doubt that sub-lattice arrangements considerably more complex than originally envisaged by Néel or Yafet and Kittel do occur in certain cases due to competing exchange interactions. A general theory has been discussed by Kaplan (1960 b), and applied to the case of Mn_3O_4 by Dwight and Menyuk (1960). From single crystal measurements they find that Jacobs' (1959 c) conclusion of a simple triangular model cannot be correct, but a more complex

PLATE I

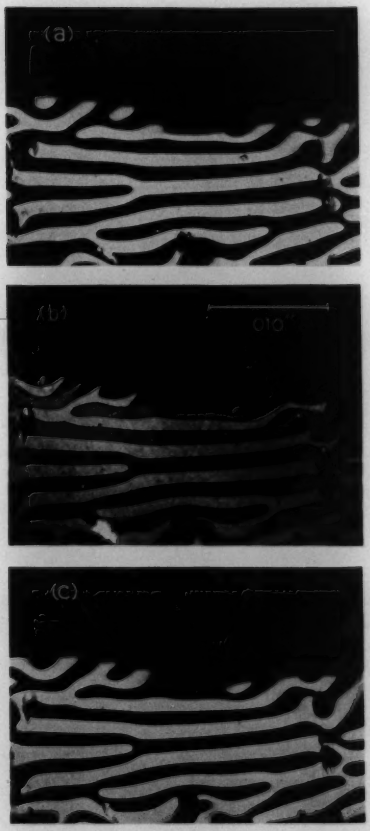


Fig. 23. An example of domains seen by transmitted polarized light in a single crystal of YIG. In Figs (*a*) and (*c*) the polarizer and analyser are crossed by similar amounts but in opposite senses ; in (*b*) they are parallel and the regions between the domains (the Bloch walls) become visible. The upper half of the field of view shows a domain with magnetization perpendicular to the line of observation (*after* Dillon and Earl 1959 b).

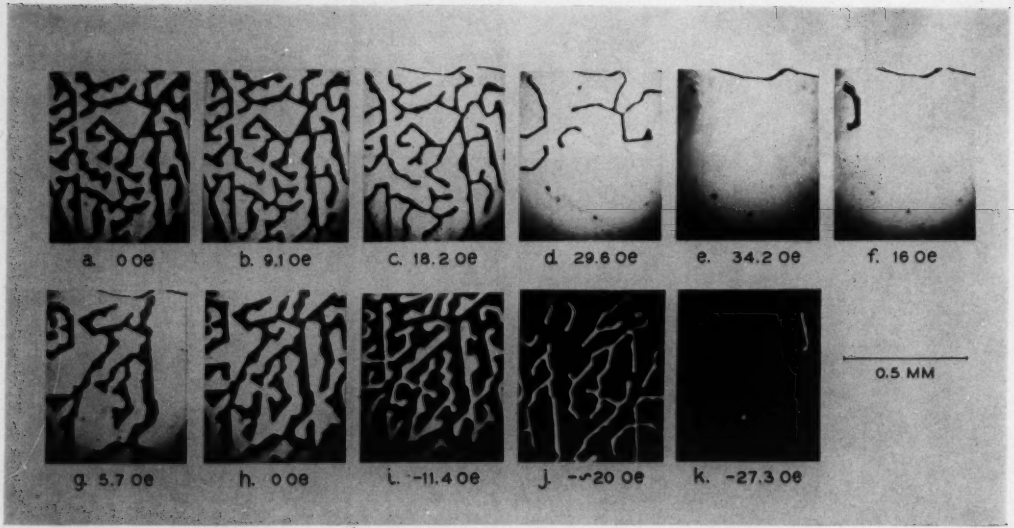


Fig. 24. Domains in Gd1G with a variety of axial fields. This compound is used since it has a very low magnetization at room temperature and a thin section can be saturated with easily available fields, even normally to its plane. These photographs were taken in sequence with no change other than that of the field. The field was run to about 70 Oe between (*e*) and (*f*). Note the hysteresis (compare (*c*) and (*f*)) (*after* Dillon 1958 b).

PLATE II

canted spin arrangement accounts for the data. One of the main difficulties, both experimental and theoretical, seems to lie in allowing adequately for anisotropy effects which are important in many real substances.

One case in which the sub-lattice arrangement is dominated completely by the effect of local anisotropy has recently been reported by Herpin, Koehler and Mériel (1960). From neutron diffraction measurements on powdered holmium iron garnet, they found that this substance has effectively 8 magnetic sub-lattices. As in the case of YIG the iron moments form two antiparallel sub-lattices, with a preferential direction along one of the [111] axes. Relative to this particular [111] direction there are six types of Ho ions, with moments given by

$$(\mu_1)_x = 7{\cdot}16\mu_B, \quad (\mu_1)_y = 2{\cdot}48\mu_B, \qquad (\mu_1)_z = 2{\cdot}48\mu_B;$$
$$(\mu_2)_x = 7{\cdot}96\mu_B, \quad (\mu_2)_y = -0{\cdot}34\mu_B, \quad (\mu_2)_z = -0{\cdot}34\mu_B,$$

together with cyclic permutation of x, y, z (corresponding to trigonal symmetry about the [111] axis). This arrangement may be described alternatively as being composed of six Ho sub-lattices, each with ions of magnitude $8\mu_B$, the moments in three of them making angles of 29° and in the other three angles of 63° with the resultant of two strictly antiparallel Fe sub-lattices.

8.2. *Sub-lattice Magnetization*

A very elegant determination of the temperature variation of the sub-lattice magnetizations in YIG has been reported by Robert (1960). Using the results of Mössbauer γ-ray absorption experiments reported earlier by Solomon (1960, see also Alff and Wertheim 1960), he was able to find the nuclear magnetic resonance for ^{57}Fe in both the [*a*] and (*d*) sites, and follow the variation of the frequency with temperature. The range covered was 1·7° to 400°K. Assuming that the hyperfine structure is given by $A\mathbf{S}.\mathbf{I}$, with A independent of temperature, the nuclear magnetic resonance frequency is given by $A\langle S\rangle$, and its variation therefore gives an accurately measurable determination of the temperature dependence of the average value of $\mathbf{S}$ on the particular sub-lattice. The curves found for the two sub-lattices were not in very close agreement with the values deduced from magnetic moments by Pauthenet (1958), but the converse procedure, that of calculating the temperature variation of the total magnetization, is in good agreement with the experimental values. It should be possible to repeat this experiment in many other ferrimagnetic substances containing ^{57}Fe, and it may be hoped that accurately determined sub-lattice magnetizations will ultimately stimulate the development of statistical theories superior to those in use at present.

In an important paper on this subject, of which only an abstract has appeared so far, Walker (1961) has pointed out that it is necessary to consider zero-point fluctuations in ferrimagnetic systems, just as in the antiferromagnetic case (see for example Nagamiya, Yosida and Kubo 1955). This leads to some spin disorder even at $T = 0$°K and arises from the fact that, in contrast to the ferromagnetic case, a completely aligned arrangement is not an eigenstate of the exchange Hamiltonian (Eqn (2.1)) when $J > 0$. One consequence of this residual disorder would be an appreciable reduction of the crystalline anisotropy from the value calculated on the basis of a molecular field, since this is very sensitive to the degree of saturation of

the individual sub-lattices. It is not yet quite clear whether the effect of the zero-point motion is observable in the magnitude of the resultant magnetization, but it appears that the effect in different sub-lattices just cancels.

8.3. *Spin Waves and Non-linear Effects*

Following the experiments of Schlömann, Green and Milano (1960), in which spin-waves were generated by the simultaneous application of *parallel* d.c. and radio-frequency fields (see § 5.8), Turner (1960) has recently observed some striking critical field effects in YIG. Working at a fixed frequency (34·6 Gc/s) and varying the applied d.c. field, he found two abrupt changes in the microwave power level required for the onset of 'saturation', and he interpreted these as due to direct interaction between the generated spin-waves and phonons (transverse and longitudinal) having the same wave number k and energy $\hbar\omega$. For phonons $k \propto \omega$ (through the velocity of sound), and by varying the frequency and d.c. field together it should therefore be possible to generate spin-waves having any specific value of k. This was confirmed by further measurements at 11·4 Gc/s. From an analysis of the experiments it was possible to estimate the exchange stiffness constant D (see Eqn (4.1)), and reasonable agreement with the specific heat measurements was found. Extending the experiments over a wide range of temperatures LeCraw and Walker (1961) were able to investigate the variation of D from 4° to 450°K. The theory of spin-wave phonon interaction has recently been discussed by Schlömann (1960 b), who also refers to earlier work.

In addition to the papers mentioned above, there have been a considerable number of other developments which were reported briefly at the New York Magnetism Conference, held in November 1960. Altogether some 43 papers dealing with ferrimagnetism and related topics were presented, and it is clearly impossible to review all these here. The main emphasis of current interest seems to be centred on the topics under the headings (8.1) and (8.3) above. There were also several papers dealing with improved methods of preparing materials, notably one by Albernethy, Ramsey and Ross describing a zone melting method of growing single crystals of YIG. There was also an important group of papers dealing with theories and experimental determinations of crystalline anisotropy. For details of all these communications reference must be made to the Proceedings, which are due to be published in a special issue of the *Journal of Applied Physics* in March 1961.

Acknowledgments

The author would like to thank many colleagues for helpful and stimulating discussions. In particular he wishes to express his gratitude to Professor J. H. Van Vleck, Professor C. L. Hogan and other members of Harvard University for introducing him to many aspects of this field, to Dr. A. H. Cooke for suggesting and encouraging the writing of this report, to Dr. D. T. Edmonds, Dr. I. S. Jacobs, Professor R. V. Jones, Dr. M. J. M. Leask and Dr. J. C. Slonczewski for helpful comments and to Dr. E. A. Wolf for valuable help during the preparation of the manuscript. This work was started during the tenure of an I.C.I. Research Fellowship, and it was supported in part by the United States Air Force.

Permission from authors and editors to reproduce the figures used in this report is gratefully acknowledged.

References

Abragam, A., and Pryce, M. H. L., 1951, *Proc. Roy. Soc.* A, **205**, 135.

Adelsköld, V., 1938, *Ark. Kemi. Min. Geol.*, **12A**, 1.

Akhieser, A., 1946, *J. Phys. U.S.S.R.*, **10**, 217.

Aléonard, R., 1960, *J. Phys. Chem. Solids*, **15**, 167.

Aléonard, R., and Barbier, J.-C., 1957, *C.R. Acad. Sci., Paris*, **245**, 831 ; 1959, *J. Phys. Radium*, **20**, 378.

Aléonard, R., Barbier, J.-C., and Pauthenet, R., 1956, *C.R. Acad. Sci., Paris*, **242**, 2531.

Alff, C., and Wertheim, G. K., 1960, *Bull. Amer. Phys. Soc.*, [II], **5**, 428.

Anderson, P. W., 1950, *Phys. Rev.*, **79**, 350; 1952, *Phys. Rev.*, **86**, 694; 1956, *Phys. Rev.*, **102**, 1008 ; 1959, *Phys. Rev.*, **115**, 2.

Anderson, P. W., and Suhl, H., 1955, *Phys. Rev.*, **100**, 1788.

Artman, J. O., 1956, *Proc. Inst. Radio Engrs, N.Y.*, **44**, 1284 ; 1957 a, *Phys. Rev.*, **105**, 62 ; 1957 b, *J. Appl. Phys.*, **28**, 92.

Artman, J. O., and Tannenwald, P. E., 1953, *Phys. Rev.*, **91**, 1014 ; 1955, *J. Appl. Phys.*, **26**, 1124.

Ayant, Y., and Thomas, J., 1959 a, *C.R. Acad. Sci., Paris*, **248**, 387 ; 1959 b, *C.R. Acad. Sci., Paris*, **248**, 1955 ; 1960, *C.R. Acad. Sci., Paris*, **250**, 2688.

Bacon, G. E., 1955, *Neutron Diffraction* (Oxford : University Press).

Bacon, G. E., and Roberts, F. F., 1953, *Acta Cryst.*, **6**, 57.

Bagguley, D. M. S., and Owen, J., 1957, *Rep. Progr. Phys.*, **20**, 304.

Bailey, P. C., 1960, *J. Appl. Phys.*, **31**, 39(S).

Baker, J. M., and Bleaney, B., 1955, *Conference de physique des basses temperatures, Paris*, paper 83 (Paris: Institut International du Froid).

Baltzer, P. K., and Wojtowicz, P. J., 1959, *J. Appl. Phys.*, **30**, 27(S).

Barth, T. F. W., and Posnjak, E., 1932, *Z. Kristallogr.*, **82**, 325.

Bates, L. F., 1951 a, *Modern Magnetism* (Cambridge : University Press), 3rd edn ; 1951 b, *J. Phys. Radium*, **12**, 459.

Bates, L. F., and Christoffel, D. A., 1957, *Proc. Instn Elect. Engrs*, **104B**, 231.

Bates, L. F., and Marshall, G., 1953, *Rev. Mod. Phys.*, **25**, 17.

Beljers, H. G., and Polder, D., 1950, *Nature, Lond.*, **165**, 800.

Belov, K. P., and Pedko, A. V., 1960, *J. Appl. Phys.*, **31**, 55(S).

Belson, H. S., and Kriessman, C. J., 1959 a, *J. Appl. Phys.*, **30**, 170(S) ; 1959 b, *J. Appl. Phys.*, **30**, 175(S).

Berk, A. D., and Lengyel, B. A., 1955, *Proc. Inst. Radio Engrs, N.Y.*, **43**, 1587.

Bertaut, F., 1952, *C.R. Acad. Sci., Paris*, **234**, 1295 ; 1960, *C.R. Acad. Sci., Paris*, **250**, 85.

Bertaut, F., and Delorme, C., 1954, *C.R. Acad. Sci., Paris*, **239**, 504.

Bertaut, F., Deschamps, A., and Pauthenet, R., 1959, *J. Phys. Radium*, **20**, 404.

Bertaut, F., and Forrat, F., 1956 a, *C.R. Acad. Sci., Paris*, **242**, 382 ; 1956 b, *C.R. Acad Sci., Paris*, **243**, 1219 ; 1957, *C.R. Acad. Sci., Paris*, **244**, 96 ; 1958, *J. Appl. Phys.*, **29**, 247.

Bertaut, F., Forrat, F., Herpin, A., and Mériel, P., 1956, *C.R. Acad. Sci., Paris*, **243**, 898.

Bickford, L. R., Jr., 1950, *Phys. Rev.*, **78**, 449.

Bickford, L. R., Jr., Brownlow, J. M., and Penoyer, R. F., 1957, *Proc. Instn Elect. Engrs*, **104B**, 238 ; 1958, *J. Appl. Phys.*, **29**, 441.

Bickford, L. R., Jr., Pappis, J., and Stull, J. L., 1956, *Phys. Rev.*, **99**, 1210.

Birks, J. B., 1957, *Proc. Instn Elect. Engrs*, **104B**, 179.

Bleaney, B., and Stevens, K. W. H., 1953, *Rep. Progr. Phys.*, **16**, 108.

Bloch, F., 1946, *Phys. Rev.*, **70**, 460.

Bloembergen, N., 1950, *Phys. Rev.*, **78**, 572 ; 1956, *Proc. Inst. Radio Engrs, N.Y.*, **44**, 1259.

Bloembergen, N., and Wang, S., 1954, *Phys. Rev.*, **93**, 72.

Boakes, D., Garton, G., Ryan, D., and Wolf, W. P., 1959, *Proc. Phys. Soc.*, **74**, 663.

Bond, W. L., 1951, *Rev. Sci. Instrum.*, **5**, 344 ; 1954, *Rev. Sci. Instrum.*, **25**, 401.

Borovic-Romanov, A. S., and Orlova, M. P., 1957, *Soviet Phys., JETP*, **32(5)**, 1023.

Bowers, K. D., and Owen, J., 1955, *Rep. Progr. Phys.*, **18**, 304.

Bozorth, R. M., 1951, *Ferromagnetism* (New York : Van Nostrand).

Bozorth, R. M., Cetlin, B. B., Galt, J. K., Merritt, F. R., and Yager, W. A., 1955, *Phys. Rev.*, **99**, 1898.
Bozorth, R. M., and Geller, S., 1959, *J. Phys. Chem. Solids*, **11**, 263.
Bozorth, R. M., and Hamming, R. W., 1953, *Phys. Rev.*, **89**, 865.
Bozorth, R. M., Tilden, E. F., and Williams, A. J., 1955, *Phys. Rev.*, **99**, 1788.
Bozorth, R. M., and Walker, J. C., 1952, *Phys. Rev.*, **88**, 1209.
Bozorth, R. M., and Walsh, D. M., 1958, *J. Phys. Chem. Solids*, **5**, 299.
Braun, P. B., 1957, *Philips Res. Rep.*, **12**, 491.
Brockhouse, B. N., 1957, *Phys. Rev.*, **106**, 859; 1958, *Phys. Rev.*, **111**, 1273.
Brown, F., and Park, D., 1954, *Phys. Rev.*, **93**, 381.
Buehler, E., and Tanenbaum, M., 1960, *J. Appl. Phys.*, **31**, 388.
Buffler, C. R., 1959, *J. Appl. Phys.*, **30**, 172(S); see also *Thesis*, Harvard University; 1960, *J. Appl. Phys.*, **31**, 222(S).
Calhoun, B. A., 1954, *Phys. Rev.*, **94**, 1577.
Calhoun, B. A., Overmeyer, J., and Smith, W. V., 1957, *Phys. Rev.*, **107**, 993.
Callen, H. B., 1958, *J. Phys. Chem. Solids*, **4**, 256.
Callen, H. B., and Pittelli, E., 1960, *Phys. Rev.*, **119**, 1523.
Casimir, H. B. G., Smit, J., Enz, U., Fast, J. F., Wijn, H. P. J., Gorter, E. W., Duyvesteyn, A. J. W., Fast, J. D., and de Jong, J. J., 1959, *J. Phys. Radium*, **20**, 360.
Casselman, T. N., and Keffer, F., 1960, *Phys. Rev. Letters*, **4**, 498.
Chikazumi, S., 1956, *J. Phys. Soc., Japan*, **11**, 718.
Clogston, A. M., 1958, *J. Appl. Phys.*, **29**, 334; 1959, *J. Phys. Radium*, **20**, 151; 1960, *J. Appl. Phys.*, **31**, 198(S).
Clogston, A. M., Suhl, H., and Walker, L. R., 1959, *J. Phys. Chem. Solids*, **11**, 338.
Clogston, A. M., Suhl, H., Walker, L. R., and Anderson, P. W., 1956, *J. Phys. Chem. Solids*, **1**, 129.
Cloud, W. H., 1958, *Phys. Rev.*, **111**, 1046.
Cohen, J., and Ducloz, J., 1959, *J. Phys. Radium*, **20**, 402.
Coles, B. A., Orton, J., and Owen, J., 1960, *Phys. Rev. Letters*, **4**, 116.
Corliss, L. M., Hastings, J. M., and Brockmann, F. G., 1953, *Phys. Rev.*, **90**, 1013.
Damon, R. W., 1953, *Rev. Mod. Phys.*, **25**, 239.
Dillon, J. F., Jr., 1956, *Bull. Amer. Phys. Soc.*, [II], **1**, 125; 1957 a, *Phys. Rev.*, **105**, 739; 1957 b, *Bull. Amer. Phys. Soc.*, [II], **2**, 128; 1957 c, *Bull. Amer. Phys. Soc.*, [II], **2**, 238; 1958 a, *J. Appl. Phys.*, **29**, 539; 1958 b, *J. Appl. Phys.*, **29**, 1286; 1958 c, *Phys. Rev.*, **111**, 1476; 1958 d, *Phys. Rev.*, **112**, 59; 1959, *J. Phys. Radium*, **20**, 374; 1960, *J. Appl. Phys.*, **31**, 1605.
Dillon, J. F., Jr., and Earl, H. E., Jr., 1959 a, *J. Appl. Phys.*, **30**, 202; 1959 b, *Amer. J. Phys.*, **21**, 201.
Dillon, J. F., Jr., Geschwind, S., and Jaccarino, V., 1955, *Phys. Rev.*, **100**, 750.
Dillon, J. F., Jr., and Nielsen, J. W., 1959, *Phys. Rev. Letters*, **3**, 30; 1960 a, *J. Appl. Phys.*, **31**, 43(S); 1960 b, *Phys. Rev.*, **120**, 105.
Douglass, R. L., 1960, *Phys. Rev.*, **120**, 1612.
Douthett, D., and Friedberg, S. A., 1958, *Bull. Amer. Phys. Soc.*, [II], **3**, 226; see also Carnegie Inst. of Technology, Tech. Rep. 3, Nonr–760(05) NR 017–207, June 1958.
Dreyfus, B., 1955 a, *C.R. Acad. Sci., Paris*, **241**, 552; 1955 b, *C.R. Acad. Sci., Paris*, **241**, 1270; 1960, *Proceedings 7th Int. Conference on Low Temperature Physics, Toronto*, to be published.
Dunitz, J. D., and Orgel, L. E., 1957, *J. Phys. Chem. Solids*, **3**, 20, 318.
Dwight, K., and Menyuk, N., 1958, *Bull. Amer. Phys. Soc.*, [II], **3**, 41, 42; 1960, *Phys. Rev.*, **119**, 1470.
Dyson, F. J., 1956, *Phys. Rev.*, **102**, 1217, 1230.
Economos, G., 1959, *J. Amer. Ceram. Soc.*, **42**, 628.
Edmonds, D. T., and Petersen, R. G., 1959, *Phys. Rev. Letters*, **2**, 499; 1960, *Phys. Rev. Letters*, **4**, 92.
Edwards, P. L., 1959, *Phys. Rev.*, **116**, 294.
Elliott, R. J., and Lowde, R., 1955, *Proc. Roy. Soc.* A, **230**, 46.

ESCHENFELDER, A. H., 1958, *J. Appl. Phys.*, **29**, 378.
ESKOWITZ, A., and WANGSNESS, R. K., 1957, *Phys. Rev.*, **107**, 379.
FAIRWEATHER, A., WELCH, A. J. E., and ROBERTS, F. F., 1952, *Rep. Progr. Phys.*, **15**, 142.
FALLOT, M., and MARONI, P., 1951, *J. Phys. Radium*, **12**, 256.
FARRAR, R. T., 1958, *J. Appl. Phys.*, **29**, 425.
FLETCHER, P. C., and BELL, R. O., 1959, *J. Appl. Phys.*, **30**, 687.
FLETCHER, P. C., and SOLT, I. H., Jr., 1959, *J. Appl. Phys.*, **30**, 181(S); 1960, *J. Appl. Phys.*, **31**, 100(S).
FLETCHER, P. C., SOLT, I. H., Jr., and BELL, R. O., 1959, *Phys. Rev.*, **114**, 739.
FLETCHER, R. C., LECRAW, R. C., and SPENCER, E. G., 1960, *Phys. Rev.*, **117**, 955.
FOËX, G., and SERRES, A., 1950, *C.R. Acad. Sci., Paris*, **230**, 729.
FOLEN, V. J., 1960, *J. Appl. Phys.*, **31**, 166(S).
FOLEN, V. J., and RADO, G. T., 1956, *Bull. Amer. Phys. Soc.*, [II], **1**, 132; 1957, *Bull. Amer. Phys. Soc.*, [II], **2**, 263; 1958, *J. Appl. Phys.*, **29**, 438.
FONER, S., 1957, *5th Int. Conf. on Low Temperature Physics and Chemistry, Madison, Wisc.* (Madison: University of Wisconsin Press), p. 576.
FONER, S., and ARTMAN, J. O., 1958, *J. Appl. Phys.*, **29**, 443.
FORESTIER, H., and PETIT, J. C., 1960, *C.R. Acad. Sci., Paris*, **250**, 4367.
FOX, A. G., MILLER, S. E., and WEISS, M. T., 1955, *Bell System Tech. J.*, **34**, 5.
FRESH, D. L., 1956, *Proc. Inst. Radio Engrs, N.Y.*, **44**, 1303.
GALT, J. K., 1952, *Phys. Rev.*, **85**, 664; 1954, *Bell System Tech. J.*, **33**, 1023; 1957, *Proc. Instn Elect. Engrs*, **104B**, 189.
GALT, J. K., ANDRUS, J., and HOPPER, H. G., 1953, *Rev. Mod. Phys.*, **25**, 93.
GALT, J. K., MATTHIAS, B. T., and REMEIKA, J. P., 1950, *Phys. Rev.*, **79**, 391.
GALT, J. K., YAGER, W. A., and MERRITT, F. R., 1954, *Phys. Rev.*, **93**, 1119.
GELLER, S., 1960, *J. Appl. Phys.*, **31**, 30(S).
GELLER, S., BOZORTH, R. M., GILLEO, M. A., and MILLER, C. E., 1960, *J. Phys. Chem. Solids*, **12**, 111.
GELLER, S., and GILLEO, M. A., 1957 a, *Acta Cryst.*, **10**, 239, 787; 1957 b, *J. Phys. Chem. Solids*, **3**, 30.
GELLER, S., MILLER, C. E., and TREUTING, R. G., 1960, *Acta Cryst.*, **13**, 179.
DE GENNES, P.-G., 1959, *Phys. Rev. Letters*, **3**, 209.
DE GENNES, P.-G., KITTEL, C., and PORTIS, A. M., 1959, *Phys. Rev.*, **116**, 323.
GESCHWIND, S., 1959, *Phys. Rev. Letters*, **5**, 207; 1961, *Phys. Rev.*, **121**, 363.
GESCHWIND, S., and CLOGSTON, A. M., 1957, *Phys. Rev.*, **108**, 49.
GESCHWIND, S., and WALKER, L. R., 1959, *J. Appl. Phys.*, **30**, 163(S).
GILLEO, M. A., 1960, *J. Phys. Chem. Solids*, **13**, 33.
GILLEO, M. A., and GELLER, S., 1958, *Phys. Rev.*, **110**, 73; 1959, *J. Phys. Chem. Solids*, **10**, 187.
GINTSBURG, M. A., 1959, *J. Phys. Chem. Solids*, **11**, 336.
GOEDKOOP, J. A., HVOSLEF, J., and ŽIVADINOVIĆ, M., 1959, *Acta Cryst.*, **12**, 476.
GOLDMAN, J. E., 1947, *Phys. Rev.*, **72**, 529.
GOLDRING, G., SCHIEBER, M., and VAGER, Z., 1960, *J. Appl. Phys.*, **31**, 2057.
GOODENOUGH, J. B., and LOEB, A. I., 1955, *Phys. Rev.*, **98**, 391.
GORTER, E. W., 1950, *Nature, Lond.*, **165**, 798; 1954, *Philips Res. Rep.*, **9**, 295; 1957, *Proc. Instn Elect. Engrs*, **104B**, 255.
GORTER, E. W., and SCHULKES, J. A., 1953, *Phys. Rev.*, **90**, 487.
GOTTLIEB, P., 1960, *J. Appl. Phys.*, **31**, 2059.
GRIFFITHS, J. H. E., OWEN, J., PARK, J. G., and PARTRIDGE, M. F., 1959, *Proc. Roy. Soc. A*, **250**, 84.
GUILLAUD, C., and CREVEAUX, H., 1950, *C.R. Acad. Sci., Paris*, **230**, 1256.
GUILLAUD, C., and SAGE, M., 1953, *C.R. Acad. Sci., Paris*, **237**, 313.
HAMILTON, W. C., 1958, *Bull. Amer. Phys. Soc.*, [II], **3**, 42.
HARRISON, F. W., OSMOND, W. P., and TEALE, R. W., 1957, *Phys. Rev.*, **106**, 856.
HASTINGS, J. M., and CORLISS, L. M., 1953, *Rev. Mod. Phys.*, **25**, 114; 1956, *Phys. Rev.*, **104**, 328.

HAUSER, W., and BROWN, L., 1959, *J. Appl. Phys.*, **30**, 1460.
HEALY, D., 1952, *Phys. Rev.*, **86**, 1009.
HELLER, G., and KRAMERS, H. A., 1934, *Proc. Acad. Sci. Amsterdam*, **37**, 378.
HERPIN, A., KOEHLER, W. C., and MÉRIEL, P., 1960, *C.R. Acad. Sci., Paris*, **251**, 1359.
HERRING, C., and KITTEL, C., 1951, *Phys. Rev.*, **81**, 869.
HIRAHARA, E., and MURAKAMI, M., 1958, *J. Phys. Chem. Solids*, **7**, 281.
HOGAN, C. L., 1952, *Bell System Tech. J.*, **31**, 1 ; 1953, *Rev. Mod. Phys.*, **25**, 253 ; 1956, *Proc. Inst. Radio Engrs, N.Y.*, **44**, 1345.
HOLSTEIN, T., and PRIMAKOFF, H., 1940, *Phys. Rev.*, **58**, 1098 ; 1941, *Phys. Rev.*, **59**, 388.
HURD, R. A., 1958, *Canad. J. Phys.*, **36**, 1072.
IIDA, S., 1960, *J. Appl. Phys.*, **31**, 251(S).
IIDA, S., SEKIZAWA, H., and AIYAMA, Y., 1955, *J. Phys. Soc., Japan*, **10**, 907.
INOUE, T., and IIDA, S., 1958, *J. Phys. Soc., Japan*, **13**, 656.
INOUE, T., MIZUTA, H., and IIDA, S., 1960, *J. Phys. Soc. Japan*, **15**, 1899.
JACOBS, I. S., 1959 a, *J. Appl. Phys.*, **30**, 301(S) ; 1959 b, *Bull. Amer. Phys. Soc.*, [II], **4**, 178 ; 1959 c, *J. Phys. Chem. Solids*, **11**, 1 ; 1960, *J. Phys. Chem. Solids*, **15**, 54.
JACOBS, I. S., and LAWRENCE, P. E., 1958, *Rev. Sci. Instrum.*, **29**, 713.
JARRET, H. S., and WARING, R. K., 1958, *Phys. Rev.*, **111**, 1223.
JONES, R. V., RODRIGUE, G. P., and WOLF, W. P., 1958, *J. Appl. Phys.*, **29**, 434.
JONKER, G. H., 1959, *J. Phys. Chem. Solids*, **9**, 165.
JONKER, G. H., WIJN, H. P. J., and BRAUN, P. B., 1956, *Philips Tech. Rev.*, **18**, 145 ; 1957, *Proc. Instn Elect. Engrs*, **104B**, 249.
KAMINOW, I. P., 1960, *Thesis*, Harvard University.
KANAMORI, J., 1959, *J. Phys. Chem. Solids*, **10**, 87 ; 1960, *J. Appl. Phys.*, **31**, 15(S).
KAPLAN, H., 1952, *Phys. Rev.*, **86**, 121.
KAPLAN, J., and KITTEL, C., 1953, *J. Chem. Phys.*, **21**, 760.
KAPLAN, T. A., 1958, *Phys. Rev.*, **109**, 782 ; 1959, *Phys. Rev.*, **116**, 888 ; 1960 a, *J. Appl. Phys.*, **31**, 364(S) ; 1960 b, *Phys. Rev.*, **119**, 1460.
KASPER, J. S., 1959, *Bull. Amer. Phys. Soc.*, [II], **4**, 178.
KASUYA, T., 1954, *Progr. Theor. Phys., Japan*, **12**, 802.
KEFFER, F., 1952, *Phys. Rev.*, **88**, 686.
KEFFER, F., KAPLAN, H., and YAFET, Y., 1953, *Amer. J. Phys.*, **21**, 250.
KEFFER, F., and KITTEL, C., 1952, *Phys. Rev.*, **85**, 329.
KITTEL, C., 1948, *Phys. Rev.*, **73**, 155 ; 1949, *Phys. Rev.*, **76**, 743 ; 1951, *Phys. Rev.*, **82**, 565 ; 1958, *Phys. Rev.*, **110**, 1295 ; 1959, *Phys. Rev.*, **115**, 1587 ; 1960, *Phys. Rev.*, **117**, 681.
KITTEL, C., and ABRAHAMS, E., 1953, *Rev. Mod. Phys.*, **25**, 233.
KNOX, K., and GELLER, S., 1958, *Phys. Rev.*, **110**, 771.
KOJIMA, Y., 1954, *Sci. Rep. Res. Inst. Tôhoku Univ.*, **A6**, 614.
KONDORSKII, E. I., PAKHOMOV, A. S., and SKIKLOSH, T., 1956, *Dokl. Akad. Nauk, S.S.S.R.*, **109**, 931 (trans. *Soviet Physics, Dokl.*, **1**, 501).
KOOY, C., 1958, *Philips Tech. Rev.*, **19**, 286.
KOUVEL, J. S., 1956, *Phys. Rev.*, **102**, 1489.
KRAMERS, H. A., 1934, *Physica*, **1**, 182.
VAN KRANENDONK, J., and VAN VLECK, J. H., 1958, *Rev. Mod. Phys.*, **30**, 1.
KRIESSMAN, C. J., HARRISON, S. E., and BLACK, S. R., 1958, *Phys. Rev.*, **110**, 844.
KRUPIČKA, S., and ZÁVĚTA, K., 1959, *Czech. J. Phys.*, **9**, 324.
KUNZLER, J. E., WALKER, L. R., and GALT, J. K., 1960, *Phys. Rev.*, **119**, 1609.
LANDAU, L., and LIFSHITZ, E., 1935, *Phys. Z. Sowjet.*, **8**, 153.
LECRAW, R. C., and SPENCER, E. G., 1959, *J. Appl. Phys.*, **30**, 185(S) ; 1960, *Phys. Rev. Letters*, **4**, 130.
LECRAW, R. C., SPENCER, E. G., and PORTER, C. S., 1958 a, *J. Appl. Phys.*, **29**, 326 ; 1958 b, *Phys. Rev.*, **110**, 1311.
LECRAW, R. C., and WALKER, L. R., 1961, *Proceedings of 6th Conference on Magnetism and Magnetic Materials, New York* 1960, *J. Appl. Phys.*, **32**, 167(S).
LI, Y. Y., 1951, *Phys. Rev.*, **84**, 721.
LOTGERING, F. K., 1956 a, *Philips Res. Rep.*, **11**, 190 ; 1956 b, *Philips Res. Rep.*, **11**, 218.

McClure, D. S., 1957, *J. Phys. Chem. Solids*, **3**, 311.
McGuire, T. R., 1955 a, *Phys. Rev.*, **97**, 831 ; 1955 b, *Proc. Amer. Inst. Elect. Engrs Conf. on Magnetism and Magnetic Materials, Pittsburgh*, T-78, p. 43.
McGuire, T. R., Howard, L. N., and Smart, J. S., 1952, *Ceramic Age*, **60**, 22.
Mason, B., 1947, *Amer. Mineralogist*, **32**, 426.
Masters, J. I., 1960, *J. Appl. Phys.*, **31**, 41(S).
Menyuk, N., and Dwight, K., 1958, *Phys. Rev.*, **112**, 397 ; 1959, *J. Appl. Phys.*, **30**, 36(S).
Menyuk, N., Dwight, K., and Wickham, D. G., 1960, *Phys. Rev. Letters*, **4**, 119.
Menzer, G., 1929, *Z. Kristallogr.*, **69**, 300.
Mercereau, J. E., 1959, *J. Appl. Phys.*, **30**, 184(S).
Mercereau, J. E., and Feynman, R. P., 1956, *Phys. Rev.*, **104**, 63.
Merz, K. M., 1960, *J. Appl. Phys.*, **31**, 147.
Meyer, H., and Harris, A. B., 1960, *J. Appl. Phys.*, **31**, 49(S).
Miles, P. A., 1954, *Nature, Lond.*, **174**, 177 ; 1957, *Progr. Rep. XXI, Lab. for Insulation Research M.I.T.*, p. 19.
Millar, R. W., 1929, *J. Amer. Chem. Soc.*, **51**, 219.
Miller, A., 1959, *Phys. Rev.*, **116**, 1481 ; 1960, *J. Appl. Phys.*, **31**, 261(S).
Miyadai, T., Takata, H., Shichijo, Y., and Tsuya, N., 1960, *J. Phys. Soc., Japan*, **15**, 1354.
Mones, A. H., and Banks, E., 1958, *J. Phys. Chem. Solids*, **4**, 217.
Moriya, T., 1960, *Phys. Rev. Letters*, **4**, 228 ; *Phys. Rev.*, **120**, 91.
Nagamiya, T., 1953, *Progr. Theor. Phys., Japan*, **10**, 72.
Nagamiya, T., Yosida, K., and Kubo, R., 1955, *Advanc. Phys.*, **4**, 1.
Nash, W. B., and Standley, K. J., 1960, *Proc. Phys. Soc.*, **76**, 99.
Nathans, R., Pickart, S. J., Harrison, S. E., and Kriessman, C. J., 1957, *Proc. Instn Elect. Engrs*, **104B**, 217.
Néel, L., 1938, *C.R. Acad. Sci., Paris*, **206**, 49 ; 1948 a, *Ann. Phys., Paris*, **3**, 137 ; 1948 b, *J. Phys. Radium*, **9**, 193 ; 1949, *Ann. de l'Institut Fourier*, **1**, 163 ; 1951, *J. Phys. Radium*, **12**, 258 ; 1953, *Rev. Mod. Phys.*, **25**, 58 ; 1954, *J. Phys. Radium*, **15**, 225.
Néel, L., Pauthenet, R., Rimet, G., and Giron, V. S., 1960, *J. Appl. Phys.*, **31**, 27(S).
Nielsen, J. W., 1958, *J. Appl. Phys.*, **29**, 390 ; 1960, *J. Appl. Phys.*, **31**, 51(S).
Nielsen, J. W., and Dearborn, E. F., 1958, *J. Phys. Chem. Solids*, **5**, 202.
Okamura, T., and Kojima, Y., 1952, *Phys. Rev.*, **85**, 690, 693.
Okamura, T., Torizuka, Y., and Kojima, Y., 1952, *Phys. Rev.*, **88**, 1425.
Paulevé, J., 1957 a, *C.R. Acad. Sci., Paris*, **245**, 408 ; 1957 b, *C.R. Acad. Sci., Paris*, **244**, 1908 ; 1958, *J. Appl. Phys.*, **29**, 259.
Paulevé, J., and Dreyfus, B., 1956, *C.R. Acad. Sci., Paris*, **242**, 1271.
Paulevé, J., Dreyfus, B., and Soutif, M., 1959, *J. Phys. Radium*, **20**, 355.
Pauthenet, R., 1952, *Ann. Phys., Paris*, **7**, 710 ; 1958, *Ann. Chim. Phys.*, **3**, 424 ; 1959 a, *J. Appl. Phys.*, **30**, 290(S) ; 1959 b, *J. Phys. Radium*, **20**, 388.
Pearson, R. F., 1959 a, *J. Phys. Radium*, **20**, 409 ; 1959 b, *Proc. Phys. Soc.*, **74**, 505 ; 1960, *J. Appl. Phys.*, **31**, 160(S).
Pearson, R. F., and Teale, R. W., 1960, *Proc. Phys. Soc.*, **76**, 308.
Penoyer, R. F., 1956, *Proc. Amer. Inst. Elect. Engrs Conference on Magnetism, Boston*, T-91, p. 365.
Penoyer, R. F., and Bickford, L. R., Jr., 1957, *Phys. Rev.*, **108**, 271.
Pickart, S. J., and Nathans, R., 1958, *Bull. Amer. Phys. Soc.*, [II], **3**, 231 ; 1959, *Phys. Rev.*, **116**, 317.
Pippin, J. E., 1959, *Thesis*, Harvard University.
Pippin, J. E., and Hogan, C. L., 1958, *Scientific Report* No. 8, AF 19(604)-1084 Harvard University ; 1959, *Scientific Report* No. 1, AF 19(604)-1084 Harvard University.
Polder, D., 1948, *Phys. Rev.*, **73**, 1116 ; 1949, *Phil. Mag.*, **40**, 99.
Polder, D., and Smit, J., 1953, *Rev. Mod. Phys.*, **25**, 89.
Porter, C. S., Spencer, E. G., and LeCraw, R. C., 1958, *J. Appl. Phys.*, **29**, 495.
Powell, M. J. D., Gabriel, J. R., and Johnston, D. F., 1960, *Phys. Rev. Letters*, **5**, 145.
Pratt, G. W., Jr., 1955, *Phys. Rev.*, **97**, 926.
du Pré, F. K., 1955, *Philips Res. Rep.*, **10**, 1.

PRINCE, E., 1956, *Phys. Rev.*, **102**, 675 ; 1957 a, *Acta Cryst.*, **10**, 554 ; 1957 b, *Acta Cryst.*, **10**, 787.
PRINCE, E., and TREUTING, R. G., 1956, *Acta Cryst.*, **9**, 1025.
PRYCE, M. H. L., 1950, *Phys. Rev.*, **80**, 1107.
RADO, G. T., 1953, *Rev. Mod. Phys.*, **25**, 81.
RADO, G. T., and FOLEN, V. J., 1956, *Bull. Amer. Phys. Soc.*, [II], **1**, 1932 ; 1957, *Bull. Amer. Phys. Soc.*, [II], **2**, 263 ; 1960, *J. Appl. Phys.*, **31**, 62.
RADO, G. T., FOLEN, V. J., and EMERSON, W. H., 1957, *Proc. Instn Elect. Engrs*, **104B**, 198.
RADO, G. T., and TERRIS, A., 1951, *Phys. Rev.*, **83**, 177.
RADO, G. T., WRIGHT, R. W., and EMERSON, W. H., 1950, *Phys. Rev.*, **80**, 273.
RADO, G. T., WRIGHT, R. W., EMERSON, W. H., and TERRIS, A., 1952, *Phys. Rev.*, **88**, 909.
REMEIKA, J. P., 1956, *J. Amer. Chem. Soc.*, **78**, 4259 ; 1960, *J. Appl. Phys.*, **31**, 263(S).
RISTE, T., BLINOWSKI, K., and JANIK, J., 1959, *J. Phys. Chem. Solids*, **9**, 153.
ROBERT, C., 1960, *C.R. Acad. Sci., Paris*, **251**, 2685.
ROBERTS, F. F., 1951, *J. Phys. Radium*, **12**, 305.
RODRIGUE, G. P., MEYER, H., and JONES, R. V., 1960, *J. Appl. Phys.*, **31**, 376(S).
RODRIGUE, G. P., PIPPIN, J. E., WOLF, W. P., and HOGAN, C. L., 1958, *Trans. Inst. Radio Engrs*, Vol. MTT6, 83.
ROMEIJN, F. C., 1953, *Philips Res. Rep.*, **8**, 304.
ROWEN, J. H., and VON AULOCK, 1954, *Phys. Rev.*, **96**, 1151.
RUDNESS, R. G., and KEBLER, R. W., 1960, *J. Amer. Ceram. Soc.*, **43**, 17.
VAN SANTEN, J. H., and VAN WIERINGEN, J. S., 1952, *Rec. Trav. Chim.*, **71**, 420.
SATO, H., 1955, *Progr. Theor. Phys., Japan*, **13**, 119.
SATO, H., and CHANDRASEKHAR, B. S., 1957, *J. Phys. Chem. Solids*, **1**, 228.
SCHAWLOW, A. L., WOOD, D. L., and CLOGSTON, A. M., 1959, *Phys. Rev. Letters*, **3**, 271.
SCHINDLER, W. H., MCGUIRE, T. R., HOWARD, L. N., and SMART, J. S., 1952, *Phys. Rev.*, **86**, 599.
SCHLÖMANN, E., 1956, *Proc. Amer. Inst. Elect. Engrs Conf. on Magnetism and Magnetic Materials, Boston*, T-91, p. 600 ; 1958 a, *J. Phys. Chem. Solids*, **6**, 242 ; 1958 b, *J. Phys. Chem. Solids*, **6**, 257 ; 1959 a, *J. Phys. Radium*, **20**, 327 ; 1959 b, *Bull. Amer. Phys. Soc.*, [II], **4**, 53 ; 1960 a, *Phys. Rev.*, **116**, 828 ; 1960 b, *J. Appl. Phys.*, **31**, 1647.
SCHLÖMANN, E., and GREEN, J. J., 1959, *Phys. Rev. Letters*, **3**, 129.
SCHLÖMANN, E., GREEN, J. J., and MILANO, U., 1960, *J. Appl. Phys.*, **31**, 386(S).
SCHLÖMANN, E., and JONES, R. V., 1959, *J. Appl. Phys.*, **30**, 177(S).
SCHLÖMANN, E., and ZEENDER, J. R., 1958, *J. Appl. Phys.*, **29**, 341.
SCHNITZLER, A. D., FOLEN, V. J., and RADO, G. T., 1960, *J. Appl. Phys.*, **31**, 348(S).
SEAVEY, M. H., Jr., and TANNENWALD, P. E., 1958, *Phys. Rev. Letters*, **1**, 168.
SENIO, P., and TUCKER, C. W., 1953, *Rev. Sci. Instrum.*, **24**, 549.
SERRES, A., 1932, *Ann. Chim. Phys.*, [10], **17**, 5.
SHERWOOD, R. C., REMEIKA, J. P., and WILLIAMS, H. J., 1959, *J. Appl. Phys.*, **30**, 217.
SHULL, C. G., and SMART, J. S., 1949, *Phys. Rev.*, **76**, 1256.
SHULL, C. G., STRAUSER, W. A., and WOLLAN, E. O., 1951, *Phys. Rev.*, **83**, 333.
SHULL, C. G., and WOLLAN, E. O., 1956, *Solid State Physics* (New York : Academic Press), **2**, 137.
SHULL, C. G., WOLLAN, E. O., KOEHLER, W. C., 1951, *Phys. Rev.*, **84**, 912.
SLATER, J. C., 1953, *Quarterly Progress Reports*, M.I.T., July 15, 1 ; *Quarterly Progress Reports*, M.I.T., October 15, 1.
SLONCZEWSKI, J. C., 1958, *Phys. Rev.*, **110**, 1341 ; 1959, *J. Appl. Phys.*, **30**, 310(S) ; 1960, *J. Phys. Chem. Solids*, **15**, 335.
SMART, J. S., 1954, *Phys. Rev.*, **94**, 847 ; 1956, *Phys. Rev.*, **101**, 585 ; 1959, *J. Phys. Chem. Solids*, **11**, 97.
SMILTENS, J., 1952, *J. Chem. Phys.*, **20**, 990.
SMIT, J., 1959, see CASIMIR *et al.*, 1959.
SMIT, J., and BELJERS, H. G., 1955, *Philips Res. Rep.*, **10**, 113.
SMIT, J., and DUYVESTEYN, A. J. W., 1959, see CASIMIR *et al.*, 1959.

SMIT, J., LOTGERING, F. K., and ENZ, U., 1960, *J. Appl. Phys.*, **31**, 137(S).
SMIT, J., and WIJN, H. P. J., 1954, *Adv. Electronics Elect. Phys.*, **6**, 69; 1959, *Ferrites* (Eindhoven: Philips Technical Library).
SNIEDER, J., 1958, *Appl. Sci. Res.* B, **7**, 185.
SNOEK, J. L., 1947 a, *New Developments in Ferromagnetic Materials* (Amsterdam: Elsevier Publishing Co.); 1947 b, *Nature, Lond.*, **160**, 90; 1948, *Physica*, **14**, 207.
SOLOMON, I., 1960, *C.R. Acad. Sci., Paris*, **251**, 2675.
SOLT, I. H., Jr., and FLETCHER, P. C., 1960, *J. Appl. Phys.*, **31**, 101(S).
SOLT, I. H., WHITE, R. L., and MERCEREAU, J. E., 1958, *J. Appl. Phys.*, **29**, 324.
SPARKS, J. T., MEAD, W., KIRSCHBAUM, A. J., and MARSHALL, W., 1960, *J. Appl. Phys.*, **31**, 356(S).
SPARKS, M., and KITTEL, C., 1960, *Phys. Rev. Letters*, **4**, 232.
SPENCER, E. G., AULT, L. A., LECRAW, R. C., 1956, *Proc. Inst. Radio Engrs, N.Y.*, **44**, 1311.
SPENCER, E. G., and LECRAW, R. C., 1955, *J. Appl. Phys.*, **26**, 250; 1958, *Phys. Rev. Letters*, **1**, 240; 1959, *J. Appl. Phys.*, **30**, 149(S).
SPENCER, E. G., LECRAW, R. C., and CLOGSTON, A. M., 1959, *Phys. Rev. Letters*, **3**, 32.
SPENCER, E. G., LECRAW, R. C., and PORTER, C. S., 1958, *J. Appl. Phys.*, **29**, 429.
SPENCER, E. G., LECRAW, R. C., and REGGIA, F., 1956, *Proc. Inst. Radio Engrs, N.Y.*, **44**, 790.
STANDLEY, J. K., and STEVENS, K. W. H., 1956, *Proc. Phys. Soc.* B, **69**, 993.
STEVENS, K. W. H., 1953, *Rev. Mod. Phys.*, **25**, 166.
STONER, E. C., and RHODES, P., 1949, *Phil. Mag.*, **40**, 481.
SWOBODA, T. J., CLOUD, W. H., BITHER, T. A., SADLER, M. S., and JARRETT, H. S., 1960, *Phys. Rev. Letters*, **4**, 509.
SWOBODA, T. J., VAUGHAN, J. D., and TOOLE, R. C., 1958, *J. Chem. Phys. Solids*, **5**, 293.
SUHL, H., 1955, *Phys. Rev.*, **97**, 553; 1956 a, *Phys. Rev.*, **101**, 1437; 1956 b, *Proc. Inst. Radio Engrs, N.Y.*, **44**, 1270; 1957 a, *J. Phys. Chem. Solids*, **1**, 209; 1957 b, *J. Appl. Phys.*, **28**, 1225; 1957 c, *Phys. Rev.*, **106**, 384; 1958, *J. Appl. Phys.*, **29**, 416; 1959, *J. Appl. Phys.*, **30**, 1961; 1960, *J. Appl. Phys.*, **31**, 935.
TAKEI, W. J., SHIRANE, G., and FRAZER, B. C., 1960, *Phys. Rev.*, **119**, 122.
TANIGUCHI, S., and YAMAMOTO, M., 1954, *Sci. Rep. Res. Insts Tôhoku Univ.*, **A6**, 330.
TANNENWALD, P. E., 1955 a, *Phys. Rev.*, **99**, 463; 1955 b, *Phys. Rev.*, **100**, 1713; 1956, *Bull. Amer. Phys. Soc.*, [II], **1**, 126.
TAUBER, A., BANKS, E., and KEDESDY, H. H., 1958, *J. Appl. Phys.*, **29**, 385.
TOMPKINS, J. E., and SPENCER, E. G., 1957, *J. Appl. Phys.*, **28**, 969.
TSUYA, N., 1952, *Progr. Theor. Phys., Japan*, **7**, 263; 1957, *Sci. Rep. Res. Insts Tôhoku Univ.*, **B8**, 161; 1958, *J. Appl. Phys.*, **29**, 449.
TURNER, E. H., 1960, *Phys. Rev. Letters*, **5**, 100.
VALENTA, L., 1959, *Czech. J. Phys.*, **9**, 29.
VAN UITERT, L. G., 1956, *Proc. Inst. Radio Engrs, N.Y.*, **44**, 1294.
VAN VLECK, J. H., 1932, *Electric and Magnetic Susceptibilities* (Oxford: University Press); 1935, *J. Chem. Phys.*, **3**, 807; 1937, *Phys. Rev.*, **52**, 1178; 1950, *Phys. Rev.*, **78**, 266; 1951, *J. Phys. Radium*, **12**, 262; 1959, *J. Phys. Radium*, **20**, 124.
VAN VLECK, J. H., and PENNEY, W. G., 1934, *Phil. Mag.*, **7**, 961.
VERNEUIL, A., 1904, *Ann. Chim. Phys.*, [8], **3**, 20.
VERWEY, E. J. W., HAAYMAN, P. W., and ROMEIJN, F. C., 1947 *J. Chem. Phys.*, **15**, 181.
VERWEY, E. J. W., and HEILMAN, E. L., 1947, *J. Chem. Phys.*, **15**, 174.
VILLAIN, J., 1959, *J. Phys. Chem. Solids*, **11**, 303.
VILLERS, G., PAUTHENET, R., and LORIERS, J., 1959, *J. Phys. Radium*, **20**, 382.
VONSOVSKI, S. V., and SEIDOV, Y. M., 1954, *Izv. Akad. Nauk S.S.S.R., Ser. Fiz.* 3, **18**, 319.
WALDRON, R. D., 1955, *Phys. Rev.*, **99**, 1727.
WALKER, L. R., 1957, *Phys. Rev.*, **105**, 390; 1958, *J. Appl. Phys.*, **29**, 318; 1961, *Proceedings of 6th Conference on Magnetism and Magnetic Materials, New York* 1960, *J. Appl. Phys.*, **32**, 264(S).
WANGSNESS, R. K., 1952, *Phys. Rev.*, **86**, 146; 1953, *Phys. Rev.*, **91**, 1085; 1954, *Phys. Rev.*, **93**, 68; 1955, *Phys. Rev.*, **97**, 831; 1958, *Phys. Rev.*, **111**, 813; 1960, *Phys. Rev.*, **119**, 1496.

WATANABE, H., 1957, *Progr. Theor. Phys.* (*Kyoto*), **18**, 405.
WEIL, L., BERTAUT, F., and BOCHIROL, L., 1950, *J. Phys. Radium*, **11**, 208.
WEISS, M. T., 1957, *Phys. Rev.*, **107**, 317 ; 1958, *Phys. Rev. Letters*, **1**, 239 ; 1959, *J. Appl. Phys.*, **30**, 146(S) ; 1960, *J. Appl. Phys.*, **31**, 778, 103(S).
WEISS, M. T., and ANDERSON, P. W., 1955, *Phys. Rev.*, **98**, 925.
WEISS, M. T., and FOX, A. G., 1952, *Phys. Rev.*, **88**, 146.
WEISS, P., and FORRER, R., 1926, *Ann. Phys., Lpz.*, **5**, 153.
WEISS, P. R., 1948, *Phys. Rev.*, **74**, 1493.
WENCUS, J. F., and LEAVITT, W. Z., 1956, *Proc. Amer. Inst. Elect. Engrs Conf. on Magnetism and Magnetic Materials, Boston*, T-91, p. 526.
WENT, J. J., RATHENAU, G. W., GORTER, E. W., and VAN OOSTERHOUT, G. W., 1952, *Philips Tech. Rev.*, **13**, 194.
WENT, J. J., and WIJN, H. P. J., 1951, *Phys. Rev.*, **82**, 269.
WHITE, D. L., 1960, *Bull. Amer. Phys. Soc.*, [II], **5**, 189.
WHITE, R. L., 1959 a, *J. Appl. Phys.*, **30**, 182(S) ; 1959 b, *Phys. Rev. Letters*, **2**, 465 ; 1959 c, *Phys. Rev.*, **115**, 1519 ; 1960, *J. Appl. Phys.*, **31**, 86(S).
WHITE, R. L., and ANDELIN, J. P., Jr., 1959, *Phys. Rev.*, **115**, 1435.
WHITE, R. L., and CARSON, J. W., 1960, *J. Appl. Phys.*, **31**, 53(S).
WHITE, R. L., and SOLT, I. H., Jr. 1956, *Phys. Rev.*, **104**, 56 ; 1957, *Bull. Amer. Phys. Soc.*, [II], **2**, 22.
WICKERSHEIM, K. A., LEFEVER, R. A., and HANKING, B. M., 1960, *J. Chem. Phys.*, **32**, 271.
WICKERSHEIM, K. A., and WHITE, R. L., 1960, *Phys. Rev. Letters*, **4**, 123.
WICKHAM, D. G., 1959, *M.I.T. Lincoln Lab. Report*, 53–7.
WICKHAM, D. G., and CROFT, W. J., 1958, *J. Phys. Chem. Solids*, **7**, 351.
WICKHAM, D. G., and GOODENOUGH, J. B., 1959, *Phys. Rev.*, **115**, 1156.
VAN WIERINGEN, J. S., 1953, *Phys. Rev.*, **90**, 488.
WIJN, H. P. J., 1952, *Nature, Lond.*, **170**, 707.
WIJN, H. P. J., GEVERS, M., and VAN DER BURGT, C. M., 1953, *Rev. Mod. Phys.*, **25**, 91.
WIJN, H. P. J., and WENT, J. J., 1951, *Physica*, **17**, 976.
WILLIAMS, H. J., BOZORTH, R. M., and GOERTZ, M., 1953, *Phys. Rev.*, **91**, 1107.
WILLIAMS, H. J., HEIDENREICH, R. D., and NESBITT, E. A., 1956, *J. Appl. Phys.*, **27**, 85.
WOJTOWICZ, P. J., 1959, *Phys. Rev.*, **116**, 32 ; 1960, *J. Appl. Phys.*, **31**, 265(S).
WOLF, W. P., 1957, *Phys. Rev.*, **108**, 1152 ; 1959, *Proc. Phys. Soc.*, **74**, 665 ; 1960, *J. Phys. Soc., Japan*, **15**, 2104 ; 1961, *J. Appl. Phys.* **32**, in the press.
WOLF, W. P., and RODRIGUE, G. P., 1958, *J. Appl. Phys.*, **29**, 105.
WOLF, W. P., and VAN VLECK, J. H., 1960, *Phys. Rev.*, **118**, 1490.
WOOD, D. L., 1959 (unpublished, quoted by SHERWOOD *et al.*, 1959).
YAFET, Y., and KITTEL, C., 1952, *Phys. Rev.*, **87**, 290.
YAGER, W. A., GALT, J. K., and MERRITT, F. R., 1955, *Phys. Rev.*, **99**, 1203.
YAGER, W. A., GALT, J. K., MERRITT, F. R., and WOOD, E. A., 1950, *Phys. Rev.*, **80**, 744.
YAGER, W. A., MERRITT, F. R., and GUILLAUD, C., 1951, *Phys. Rev.*, **81**, 477.
YASUCOCHI, K., KANEMATSU, K., and OHOYAMA, T., 1960, *J. Phys. Soc., Japan*, **15**, 932.
YODER, H. S., and KEITH, M. L., 1951, *Amer. Min.*, **36**, 519.
YOSHIMORI, A., 1959, *J. Phys. Soc., Japan*, **14**, 807.
YOSIDA, K., and TACHIKI, M., 1957, *Progr. Theor. Phys., Japan*, **17**, 331.
ZÁVĚTA, K. (reported by VALENTA, L.), 1959, *J. Phys. Radium*, **20**, 414.
ZEENDER, R. J., and SCHLÖMANN, E., 1960, *J. Appl. Phys.*, **31**, 1112.

Technical Applications

FOX, A. G., MILLER, S. E., and WEISS, M. T., 1955, *Bell. Syst. Tech. J.*, **34**, 5.
HOGAN, C. L., 1956, *Proc. Inst. Radio Engrs, N.Y.*, **44**, 1345.
LAX, B., 1956, *Proc. Inst. Radio Engrs, N.Y.*, **44**, 1369.
OWENS, C. D., 1956, *Proc. Inst. Radio Engrs, N.Y.*, **44**, 1234.
SOOHOO, R. F., 1960, *Theory and Application of Ferrites* (New Jersey : Prentice Hall).
WEISS, M. T., 1957, *Phys. Rev.*, **107**, 317.

See also many articles in the following Conference Proceedings :

Ferrites, 1956, London, *Proc. Instn Elect. Engrs*, **104B**, (Suppl.) pp. 127–576 (1957).

Magnetism and Magnetic Materials, 1952, University of Maryland (*Rev. Mod. Phys.*, **25**, 1–351) ; *Magnetism and Magnetic Materials*, 1955, Pittsburgh (*Amer. Inst. Elect. Engrs*, T-78) ; *Magnetism and Magnetic Materials*, 1956, Boston (*Amer. Inst. Elect. Engrs*, T-91) ; *Magnetism and Magnetic Materials*, 1957, Washington (*J. Appl. Phys.*, **29**, 237–548, 1958) ; *Magnetism and Magnetic Materials*, Philadelphia (*J. Appl. Phys.*, **30**, (Suppl.), 1959) ; *Magnetism and Magnetic Materials*, 1959, Detroit (*J. Appl. Phys.*, **31**, (Suppl.), 1960).

Properties and Applications of Ferrites, Harvard University, 1956, *Proc. Inst. Radio Engrs, N.Y.*, **44**, 1234–1468.

THEORY AND APPLICATIONS OF THE DENSITY MATRIX

By D. TER HAAR
The Clarendon Laboratory, Oxford

CONTENTS

	PAGE
§ 1. Introduction	304
§ 2. General properties of the density matrix	312
§ 3. Green function techniques	319
§ 4. The description of statistical equilibrium by the density matrix	322
§ 5. Density matrix techniques applied to atoms, molecules and nuclei	327
§ 6. The density matrix in solid state physics	334
§ 7. Non-equilibrium processes; transport theory	337
§ 8. Polarization, scattering, and angular correlation experiments	341
§ 9. Resonance and relaxation phenomena	347
§ 10. The theory of measurement	353
Concluding remarks and acknowledgments	356
References	356

Abstract. After a qualitative discussion of the advantages of the density matrix and of the different ways to introduce it (the statistical, quantum mechanical and operational methods of approach), § 2 deals with the general properties of the density matrix, including a discussion of pure cases and mixtures. A brief discussion is given of Green function techniques and of the relation between Green functions and correlation functions. A discussion of recent developments in the evaluation of partition functions concludes the first part of this article dealing with the theory of density matrix techniques. Sections 5 to 9 discuss applications. The first application is the quantum-chemical one to many-body systems in their ground state, that is, systems at absolute zero, and it is shown how the density matrix fits into the Hartree–Fock and Thomas–Fermi schemes. A brief discussion is given of the theory of diamagnetism. This is followed by a discussion of non-equilibrium processes and of Kubo's approach to transport theory. After that the polarization of beams of electrons or of photons is discussed and it is indicated how density matrix techniques can be used to treat scattering processes. Section 9 concludes this part of the paper by a brief account of density matrix theory applications to resonance and relaxation phenomena. Finally, the theory of measurement in quantum mechanics is considered.

§ 1. Introduction

The recent rapid growth of physics is producing not only an ever-growing spate of papers—often repeating what has been published independently elsewhere—but also an increasing degree of specialization. Different branches of physics often use different terms for the same concept, and it takes a long time before new techniques filter through from one branch of physics to another. This is a great pity since in theoretical physics, for instance, techniques which are useful in one field are often also applicable to other fields. Two such techniques will be discussed in the present paper. The first—and main—one is the use of the so-called density matrix which was introduced by von Neumann (1927 a, b, c; see also Dirac 1929, 1935, and von Neumann 1932), and was until recently used mainly in statistical mechanics. The other one is the use of the so-called Green functions which were originally used in field theory and which

have recently been applied with great success to a great number of many-body problems (for an extensive survey of Green function techniques applied to statistical problems we refer to a paper by Zubarev (1960)). In this introduction we shall first of all discuss why the density matrix was introduced and how its use fits in easily with a description of different kinds of experiments. We shall then give a brief outline of the subsequent sections. The aim of this introduction is that anybody who wants to get acquainted with the use of the density matrix in various branches of physics without being interested in details should get sufficient information from the introduction. Those readers who wish to apply density matrix techniques themselves and would thus be interested in a more detailed account are referred to the sections dealing with the various applications. There is also an extensive list of references, both of those quoted in the text and of some related papers which are not referred to explicitly.

The density matrix was introduced by von Neumann to describe statistical concepts in quantum mechanics. To see how this can be done we must discuss the fundamental ideas both of quantum mechanics and of statistical mechanics. Most physical systems consist of so many particles, or possess so many degrees of freedom, that it is impossible to specify completely the state of these systems. Indeed, physicists are forced to make predictions about the behaviour of the systems they study from a knowledge of a very small number of parameters. That this is possible, and that those predictions in general will be in accordance with the actual behaviour of physical systems, is because one can use statistical methods and introduce so-called *representative ensembles* (see, for instance, Tolman 1938, ter Haar 1954, 1955, for a discussion of this point of view). Such representative ensembles are collections of identical systems such that the values of those parameters which are known are the same for all systems in the ensemble while the other parameters will have values which are spread over a suitably chosen range for the different systems. The methods by which such representative ensembles are chosen are discussed in statistical mechanics, and we shall briefly return to possible choices and the reasons why they are made in § 4. Their success arises because it turns out that the average value of any physical quantity, the average being taken in the appropriate manner over the ensemble, will be equal to the actual value of that quantity for the physical system under observation—at least in as far as we may neglect fluctuations. This statistical aspect of physical systems is common to systems described by classical mechanics and those described by quantum mechanics and is due to our lack of knowledge about the details of the physical systems observed in nature.

In quantum mechanics there occurs another statistical aspect as the use of wave functions itself introduces probability predictions : the value of a physical quantity will, in general, not be well defined for a physical system in an arbitrary state, but be subject to a probability distribution. This statistical aspect remains even when we have as complete a knowledge about the physical system as is possible, that is, even when we know its wave function. In the classical case, complete knowledge would entail knowledge of all parameters describing the system and all physical quantities would thus have well-defined values. We see that statistical methods appear thus much more naturally in quantum mechanics than in classical physics. It must be emphasized, however, that for a quantum mechanical system

usually both aspects will occur. In that case it is necessary to find the quantum mechanical analogue to the classical ensemble density. As with all physical quantities in quantum mechanics, this quantum mechanical ensemble density will be an operator. It is most easily defined by its matrix elements in some appropriately chosen representation and is known as the *density matrix*. From our discussion it is clear that in quantum mechanics we would have this operator, even if we had maximum knowledge about the system considered. Such a case is called a *pure case*, and the more general case a *mixture*. The reasons for these terms will become clear presently. The density matrix can be used in the same way as the classical ensemble density to define the average values of any physical quantity, that is, one uses the normalized density matrix to construct averages. In the classical case these averages were just ordinary weighted means, while in quantum mechanics they are defined as the trace of the product of the density matrix and the operator corresponding to the physical quantity considered. In the classical case the ensemble density is determined by the number of systems which have values of the parameters defining the state of a system within given intervals. Similarly, the density matrix will be determined by stating how many systems there are in the quantum mechanical ensemble with given wave functions. The ensemble can thus in general be considered to be a superposition, or mixture, of the states corresponding to all the different wave functions represented in it. If all systems correspond to the same wave function, that is, if we know all that can be known about the system under consideration, there is no superposition, and we talk about a pure case.

Before discussing in general terms the properties and uses of the density matrix we must mention a different way of introducing the density matrix, and one where perhaps the relation between the density matrix and the lack of detailed knowledge is even more evident. It is of interest to note that Dirac in his first paper (1929; see also Dirac 1930 a) on the density matrix, where he uses the approach we have just described, never introduces the actual term density matrix. He uses this term, however, in two subsequent papers (Dirac 1930 b, 1931) where he approaches the density matrix from a slightly different angle, namely the one we are about to describe. It should be mentioned that he does not discuss in these papers the connection between the two kinds of approach and that the two kinds of density matrix—both denoted by ρ—are normalized in a different way; we shall return to this question of normalization which has led to some rather misleading statements in the literature (compare ter Haar 1960). It is also interesting to note that Landau and Lifshitz in their *Quantum Mechanics* (1958 a) use the second kind of approach, but in their *Statistical Physics* (1958 b) the first kind. The recent upsurge of density matrix techniques in quantum chemistry (see, for instance, the April 1960 issue of *Reviews of Modern Physics*, and especially Golden 1960 a, Löwdin 1960, McWeeny 1960) which will be discussed in § 5 has used practically exclusively the second kind of approach. We discussed earlier that in statistical mechanisms one introduces the density matrix essentially by defining the probability that the physical system under consideration is described by a given wave function. The situation is then characterized by the superposition, or mixture, of a large number of different wave functions. In quantum mechanics, however, one likes to work with closed systems, that is, systems which are described by a wave function,

the change in which is governed by a Hamiltonian. If we were interested in the complete and detailed behaviour of such a system, we should be dealing with a pure case and there would be no grounds for introducing the density matrix, as there would be involved only one wave function and this wave function could be used to determine the values of all physical quantities connected with the system. In general, physical systems will possess many degrees of freedom and we are often interested in the behaviour of only a few of them. To fix our ideas let us consider one of the basic problems of quantum chemistry : a many-electron system, the total energy of which we should like to evaluate. We can usually assume that our system—which may be an atom or a molecule—will be in a well-defined state, normally the ground state. If we knew the exact form of the wave function we could, of course, evaluate the energy using the known wave function and the Hamiltonian. As this is practically never the case we must have recourse to approximations. We could try to find approximations to the wave function, but we can also use the fact that the Hamiltonian of a complicated atom or molecule will contain only terms which involve the coordinates of either one or at most two electrons : the kinetic energy terms and those potential energy terms referring to nucleus–electron Coulomb energies involve one electron only, while the electron–electron interaction is a two-body Coulomb term. In evaluating the energy we are thus evaluating averages involving at most two electrons at the same time. It is therefore tempting to introduce suitable averages of the many-body wave function or rather of the product of this wave function with its adjoint over all electrons but two. Such an average will be called a *reduced* density matrix (Husimi 1940). One can prove that, indeed, these reduced one-electron and two-electron density matrices determine completely the energy of the system. Instead of looking for approximate expressions for the total many-body wave function, one can thus look for approximate expressions for the reduced density matrices.

We can generalize this approach to the density matrix as follows. Whenever we are dealing with a large system which itself is described by a wave function, we can by averaging over all degrees of freedom which are of no interest to us obtain a density matrix describing the average behaviour of those degrees of freedom over which we have not averaged. Of course, if the large system can be split into two non-interacting parts, the one corresponding to the degrees of freedom over which we average and the other corresponding to the other degrees of freedom, and if the original state of the system is described by a product wave function, the factors of which correspond to the non-interacting parts, we do not need a density matrix to describe the sub-system in which we are interested, as it possesses a wave function of its own. From this point of view we can say that the most general description of a quantum mechanical system is by a density matrix and that only in very favourable conditions is it possible to use a wave function. The extensive use of wave functions in theoretical physics is due to the fact that the description by means of a wave function is often a much simpler one.

Let us now consider the connection between the two ways of introducing density matrices, and to fix our ideas let us again consider the one-electron density matrix which we met a moment ago. In the second method of introducing the density matrix, it entered simply by averaging the product of the many-body wave function and its adjoint over all coordinates except those of one electron. In the first

method of approach we would have asked ourselves how a specially chosen electron could be described (we must emphasize that although the indistinguishability of the electrons plays an important role in determining the one-electron density matrix, or the many-body wave function, it is irrelevant to our present discussion). This would have meant that we should have constructed a representative ensemble. This ensemble would have been constructed so as to be in accordance with our knowledge about the electron, namely, the knowledge that the electron was part of an atom in a particular quantum state. The density matrix would therefore have to be constructed in such a way that the probability density function for the electron—which in fact is the diagonal element of the density matrix in coordinate representation as we shall see in the following—would be the same as the one following from the many-electron wave function of the atom, that is, the average of the absolute square of this wave function over all coordinates but those of the one electron under consideration. This means that, indeed, in coordinate representation the diagonal elements of the density matrix are uniquely defined. One can prove similarly, as we shall do in the next section, that this uniqueness extends also to the non-diagonal elements so that the density matrices defined by the two different methods are, indeed, the same. This is less surprising if we remember that a popular way of describing an ensemble, especially in classical statistical mechanics, is by assuming the different systems in the ensemble together to form a large closed system in which the original systems are interacting (see, for instance, Schrödinger 1948). The electrons in the atom can thus be considered to be the different systems, and their average behaviour is represented by the ensemble which is described by the density matrix. In using the one- or two-electron density matrices we give up all hopes of getting information about three- or more-electron correlations, that is, we sacrifice all pursuit of more detailed knowledge.

We must still discuss one more way of looking at the density matrix. So far we have seen two different ways of obtaining the density matrix, which we could call the statistical and the quantum mechanical methods of approach, and we have made it plausible that they lead to the same results. The approach used by Fano (1957) combines aspects of both methods. On the one hand it recognizes that we have often insufficiently detailed data available about the physical systems used in experiments so that we are forced to use ensembles but, on the other hand, it emphasizes that we are usually not interested in a detailed description of these systems, and that we can thus with profit use an approach which has already from the beginning eliminated superfluous degrees of freedom from the description. One might call this the operational approach. One tries to find a set of parameters which define the physical system with as much accuracy as corresponds to the experimental set-up (or the given or accessible data). The density matrix elements can often be used as such parameters and as the density matrix is often more convenient than the full quantum-mechanical description followed by the appropriate averaging process, it is used in this operational method. In the following we shall encounter examples of all three approaches. A typical example of experiments where the operational approach is suitable is that of polarization experiments (see § 8). The quantum mechanical approach is the logical one for many-body problems such as the Hartree–Fock self-consistent field or the Thomas–Fermi–Dirac

statistical theory of the atom (see § 5), and the statistical approach is the more natural one for equilibrium problems (§ 4) or transport problems (§ 7).

From this discussion we see that there is really no simple answer to the question: What is the density matrix? The answer depends completely on one's point of view and can be either: It is the quantum-mechanical counterpart of the classical distribution function (statistical point of view), or: It is the most general description of an open quantum mechanical system, that is, a system which cannot be described by a wave function (quantum-mechanical point of view), or, finally: It is the most convenient way to collect all parameters which are of interest for a given experimental set-up and to describe their behaviour (operational point of view).

In the next section we shall discuss in more detail the general properties of the density matrix, its normalization, the conditions to be satisfied for a pure state, and other conditions to be satisfied by the density matrix elements. We shall also prove the equivalence of the statistical and the quantum mechanical definitions of the density matrix and derive the equations of motion to be satisfied by the density matrix.

If one adopts the statistical point of view, one is often especially interested in equilibrium properties. The discussion of the density matrix as a tool in statistical mechanics falls to a large extent outside the scope of the present survey and can be found in most textbooks on the subject (for instance, Tolman 1938, ter Haar 1954, Kittel 1958). There have, however, been recently various developments of great importance in this subject and we wish to discuss briefly some of these developments. They have been brought about largely by the use of the new field-theoretical methods. That this is feasible can be seen as follows. In modern field theory one is concerned with the interaction of a particle with a field, or of interactions between fields. This problem can be formulated by using second quantization, that is, by putting it in terms of annihilation and creation operators. Once this is done one can use the Feynman–Dyson diagram technique to evaluate various physical quantities. In quantum statistics one is dealing with systems of interacting particles. This problem can also be expressed in terms of annihilation and creation operators, and once this is done the problem is formally of the same type as the problems encountered in field theory. A large part of the extensive literature of recent years dealing with the statistical mechanics of systems with large numbers of interacting particles has been using field theoretical methods. As a result it became clear that the field theoretical propagators or Green functions would be a useful tool. This line of attack has been developed recently, notably in Russia, especially for the treatment of many-body problems such as superconductivity and plasmas (for instance, Bonch-Bruevich 1956 a, Fradkin 1959 a, b, c, Abrikosov, Gor'kov and Dzyaloshinskiĭ 1959; see also Matsubara 1955 b, Martin and Schwinger 1959). It has also been found to be very useful in treating transport problems (see, for instance, Kubo 1957 a, Montroll and Ward 1959). It may be mentioned that Green functions were used in statistical physics before this, when they were known as time-dependent correlation functions—another instance of the difficulty of communications between one branch of theoretical physics and another. Sections 3 and 4 will be devoted to a brief discussion of this kind of problem. In § 3 we shall discuss Green function techniques and in § 4 recent

developments in the evaluation of the equilibrium density matrix of systems of interacting particles. The techniques discussed in these two sections will also be used in § 7 where we discuss transport properties and in § 6 where solid state problems are discussed.

The statistical approach to the density matrix comes to the fore in our discussions in § 4, but in a number of many-body problems such as atoms and nuclei the quantum mechanical approach is more suitable. The problems we shall discuss in § 5 are the Hartree–Fock and Thomas–Fermi–Dirac approach to atomic systems and to nuclei. We shall assume, as is usually done, that all forces are two-body forces and are additive. This is a plausible assumption in as far as electrons are concerned, but may be completely wrong for nucleons. If we do this, we can express the energy of the system in terms of one- and two-particle density matrices, and we can discuss various physical properties of the system under consideration in terms of these density matrices. As one is usually especially interested in the properties of the ground state of the system, one can use variational methods to obtain the appropriate density matrices. One can also show that the Slater determinants, normally used in Hartree–Fock theory, can be expressed in terms of these density matrices. This means that one can, in the particular case where the wave function can be expressed as a single Slater determinant, obtain even this total wave function once the single-particle density matrix is known. In the Thomas–Fermi statistical model one can use the density matrix to obtain the von Weizsäcker correction to the energy of a nucleus (Naqvi 1959). We shall also briefly discuss an ingenious method (suggested by Bopp (1959)) to use the density matrix to obtain atomic energy levels.

In §§ 6 and 7 we turn again to the statistical approach. In § 6 we discuss some equilibrium solid state problems. The density matrix is used to calculate the diamagnetism, both of a free electron gas and of the conduction electrons in metals, and a brief discussion is given of the Meissner effect in superconductors. In § 7 we discuss the density matrix approach to irreversible processes, or rather, transport processes. We shall briefly touch upon some aspects of the quantum mechanical H-theorem, that is, the problem of the approach of a quantum mechanical system to equilibrium, in § 10, but only in as far as it is involved in our discussion of the measuring process in quantum mechanics. For a more detailed discussion we must refer to the literature (see, for instance, ter Haar 1955). The discussion of irreversible processes in § 7 is restricted to the response of equilibrium systems to such external agents as electrical fields, magnetic fields, and temperature gradients. One way to treat transport properties is to use a transport equation (see, for instance, ter Haar 1954, Ch. 10, Wilson 1953). There are, however, many objections to the use of a transport equation. This is true both for classical and for quantum mechanical systems. To see the difficulties involved in the usual approach, let us fix our ideas and discuss the case of the electrical conductivity. Moreover, let us simplify the discussion by considering the classical case. To find the conductivity we must evaluate the current produced by imposing upon the system an external electrical field. This current is obtained by calculating the average velocity of the charges in the system along the direction of the electrical field. To do this we need to know the one-particle distribution function—the classical counterpart of the one-particle density matrix. This distribution function is the solution of the

Boltzmann transport equation. In this transport equation there occur the electrical field—it is usually assumed that it can be treated as a perturbation—and also a term measuring the influence of the interactions between particles (the so-called collision term). This last term will contain the two-particle distribution function. We see thus that we must either express this two-particle distribution function in terms of the one-particle distribution function, or we must first solve a transport equation for the two-particle distribution function. The first approach is usually employed in elementary transport theory but, except in very simple cases, the approximations used in simplifying the collision term are not very satisfactory. If we try to apply the second approach we run up against the difficulty that the transport equation for the two-particle distribution function contains the three-particle distribution function, and once again we face the same decision, whether to approximate here, or whether to look for the equation of motion for the three-particle distribution function. This chain is broken only when we get to the equation of motion for the n-particle distribution function (where n is the total number of particles in the system). This distribution function is nothing but the classical ensemble density, and its equation of motion is the so-called Liouville equation. We may refer at this point to the many attempts to justify the use of the Boltzmann transport equation by showing how and under what conditions it follows from the (exact) Liouville equation. Many of these attempts were described at the Brussels Conference on Transport Processes in Statistical Mechanics (Prigogine 1958). As these attempts have not been an unqualified success, many authors have used different modes of approach (Kubo 1957 a, Greenwood 1958, Kohn and Luttinger 1957, Nakano 1956, Lax 1958, Edwards 1958 b, Chester and Thellung 1959). In these papers an exact expression is found for the conductivity, or rather, an expression is obtained by approximating the exact equation of motion for the density matrix in retaining only terms which are at most linear in the field strength. We shall describe this way of evaluating transport coefficients in § 7.

The next two sections use the operational approach. The first section deals with polarization experiments, both with those dealing with particles having a magnetic moment of their own, and with those dealing with polarized light. To fix the ideas, let us consider the polarization of a beam of electrons. In this case, we are only interested in the average value of the magnetic moment of the electrons. The values of the other coordinates of the electrons are of no interest to us. There are exactly three parameters which are of interest, namely the three components of the polarization vector. One can introduce a two-by-two density matrix—which because of its normalization contains three independent parameters—to describe the system uniquely. If the beam traverses a region where there is a magnetic field, the polarization will change. To calculate this change in polarization by using the Dirac equation of an electron moving in a magnetic field and averaging over the spatial coordinates of the electron is extremely tedious. On the other hand, if one uses the equations of motion for the density matrix, and the relation between the polarization and the density matrix, the equation of motion for the polarization follows easily. It turns out to be an example of the Ehrenfest theorem which states that the equations of motion for a quantum-mechanical average will be the same as the classical equations of motion for the corresponding classical physical quantity. The polarization is the average of the magnetic moment—or

the spin—of the electrons, and its equation of motion is the classical equation for the rate of change of a magnetic moment in a magnetic field. A similar analysis can be carried out for particles with spin values larger than $\frac{1}{2}$ and also for the polarization of light.

In the last part of § 8 we discuss angular correlation and scattering experiments. Once again we are interested only in part of the parameters describing the system, and the density matrix again has turned out to be a useful tool. In § 9 we return to solid state problems, namely, the problem of the relaxation of magnetic moments in a solid. This problem is of great interest in resonance experiments. In many ways this problem—and its solution—is similar to those involved in transport processes which we discussed in § 7.

In the last section we discuss a topic of more academic interest, namely, the theory and interpretation of measurement in quantum-mechanical systems. This topic was first discussed extensively by von Neumann (1932) and is the subject of an extremely lucid book by London and Bauer (1939). It is a topic which has again been discussed recently by many authors (for instance, Ludwig 1954, Green 1958, Durand, unpublished), but we shall base our discussion on the, perhaps old-fashioned, point of view presented by London and Bauer. The main aspects of the measuring process are the following ones. The system to be observed must be coupled with a set of detectors. A definite change in the state of the detectors will be brought about by this coupling. The measuring set-up will depend upon the physical quantity to be measured and will be such that different values of the physical quantity involved will lead to different detectors being activated. After the interaction between the system and the detector the two are decoupled, and from seeing which detector has been activated we can conclude the value of the physical quantity which corresponded to the original state of the system under observation. The measurement itself has, however, changed the state of the observed system. Before the measuring process the system might have been in a pure state, but this will no longer be the case after the measurement.

In this introduction we have attempted to give a rough idea of the problems connected with the density matrix. In the limited confines of a survey article one must of necessity restrict the discussion and in the next sections we shall try to give a more quantitative discussion of the topics mentioned in the foregoing, but we shall still have to leave out many details for which we must refer to the literature.

§ 2. General Properties of the Density Matrix

In this section we shall discuss the general properties of the density matrix. We shall introduce the density matrix first of all through the statistical approach and prove some of its properties. After that we shall introduce it through the quantum-mechanical approach and show that the two density matrices are, indeed, the same. We may refer to the following papers and books to supplement the discussion in this section : Kemble 1937, Tolman 1938, Husimi 1940, ter Haar 1954, Fano 1957 and Hagedorn 1958.

In the statistical approach to the density matrix we are concerned with the description of a physical system by an ensemble. This ensemble will often be a

grand ensemble, that is, the number of particles in the different constituent systems will not necessarily be the same for all systems. We shall, however, in this section not emphasize this aspect, but it will be important in the discussion in § 4. Let there be N systems in the ensemble and let these systems be described by normalized wave functions ψ^k $(k = 1, \ldots, N)$. It is convenient to introduce a complete orthonormal set φ_n in terms of which the ψ^k can be expanded. To simplify our discussion in the present section we shall neglect complications introduced by spin, that is, we assume the ψ^k and φ_n to be scalars. It is easy enough to take spin into account, and we shall do this, for instance, in § 8. In terms of the φ_n we have

$$\psi^k = \sum_n c_n{}^k \varphi_n, \quad (k = 1, \ldots, N). \qquad \ldots\ldots(2.1)$$

Originally we had defined our ensemble by the wave functions ψ^k. In the new representation it is defined by the coefficients $c_n{}^k$.

Let A be a physical quantity, the value of which we wish to determine for the system which is represented by our ensemble. This quantity will correspond quantum mechanically to an operator $\hat{A}$ (all operators will be denoted by $\hat{}$) and its average value for the kth system will be given by the equation

$$A_k = \int \psi^{k*}\, \hat{A} \psi^k \, d\tau, \qquad \ldots\ldots(2.2)$$

where $\int \ldots d\tau$ indicates integration over all arguments of ψ^k. Taking the average over the ensemble, which we shall denote by $\langle \ldots \rangle$, we obtain the expectation value of $\hat{A}$ for the system under consideration, and we get

$$\langle \hat{A} \rangle = N^{-1} \sum_{k=1}^{N} \int \psi^{k*}\, \hat{A} \psi^k \, d\tau, \qquad \ldots\ldots(2.3)$$

or, using Eqn (2.1),

$$\langle \hat{A} \rangle = N^{-1} \textstyle\sum_k \sum_{m,n} c_m{}^{k*}\, c_n{}^k \, A_{mn}, \qquad \ldots\ldots(2.4)$$

where A_{mn} is the matrix element of $\hat{A}$ in the φ_n-representation :

$$A_{mn} = \int \varphi_m{}^*\, \hat{A} \varphi_n \, d\tau. \qquad \ldots\ldots(2.5)$$

Introducing the density matrix $\hat{\rho}$ by its matrix elements in the φ_n-representation,

$$\rho_{mn} = N^{-1} \textstyle\sum_k c_n{}^{k*}\, c_m{}^k, \qquad \ldots\ldots(2.6)$$

we can write Eqn (2.4) in the form

$$\langle \hat{A} \rangle = \textstyle\sum_{m,n} A_{mn}\, \rho_{nm} = \mathrm{Tr}\,(\hat{\rho}\hat{A}), \qquad \ldots\ldots(2.7)$$

where Tr indicates the trace, that is, the sum of the diagonal elements.

A few remarks should be made here about Eqns (2.3), (2.6) and (2.7). First of all we notice that $\langle \hat{A} \rangle$ is a double average : one average is the quantum-mechanical average given by Eqn (2.2), and the other is the statistical average over the ensemble. This means that two different kinds of probability considerations enter into our discussion ; this point was specially emphasized by London and Bauer (1939). The quantum-mechanical probability considerations enter even if we have as complete a knowledge about the physical system as is possible in quantum mechanics, that is, if we know the wave function of the system ; they are a feature inherent in quantum mechanics and are not related to any lack of (possible) knowledge. The

statistical probability considerations, on the other hand, are closely connected to our lack of knowledge—as they are in classical statistics—and were introduced exactly because our knowledge is incomplete. Secondly, we note that the density matrix, or *statistical operator* as it is sometimes called, is defined by its matrix elements in the particular representation in which we are working. We shall see presently that if we change to another representation the density matrix will change according to the usual rules of quantum mechanical transformation theory. Finally, we note the intimate connection between the density matrix elements and the average values of physical quantities. This connection is used by the operational approach to define the density matrix in terms of averages. There are cases where the density matrix is a finite matrix—for instance, when we are dealing with polarization experiments—and if it is a matrix of rank M it is determined by M^2-1 independent parameters (*vide infra*). If we can find the average values of M^2-1 physical quantities A_j these will suffice to determine the density matrix. In the case of the polarization of light, for instance, M will be 2, and the three components of the polarization vector will just be sufficient to determine the density matrix completely, as we shall see in § 8.

From the definition (2.6) we see that $\hat{\rho}$ is Hermitian,

$$\rho_{mn} = \rho_{nm}^*. \qquad \ldots\ldots(2.8)$$

Applying Eqn (2.7) to the unit operator ($\hat{A} = \hat{1}$) we find that $\hat{\rho}$ is normalized (to unity),

$$1 = \langle \hat{1} \rangle = \mathrm{Tr}\, \hat{\rho}.\hat{1} = \mathrm{Tr}\, \hat{\rho}. \qquad \ldots\ldots(2.9)$$

This result could also have been obtained directly from Eqn (2.6) and the fact that the ψ^k are normalized

$$\mathrm{Tr}\, \hat{\rho} = N^{-1} \sum_k \sum_n a_n{}^{k*} a_n{}^k = N^{-1} \sum_k 1 = 1. \qquad \ldots\ldots(2.10)$$

From Eqn (2.8) it follows that the diagonal elements of the density matrix, ρ_{nn}, are real, and from Eqn (2.10) it follows that they must satisfy the relations

$$\sum_n \rho_{nn} = 1, \quad 0 \leqslant \rho_{nn} \leqslant 1. \qquad \ldots\ldots(2.11)$$

The physical meaning of the ρ_{nn} is clear from Eqns (2.6) and (2.1): it is the (normalized) probability that φ_n is realized in the ensemble.

Let us now consider a change from the φ_n-representation to another representation, say, the χ_p-representation. Instead of Eqn (2.1) we have now

$$\psi^k = \sum_n c_n{}^k \varphi_n = \sum_p d_p{}^k \chi_p. \qquad \ldots\ldots(2.12)$$

The transformation from the φ_n- to the χ_p-representation will be characterized by a unitary transformation matrix S_{np} such that

$$\chi_p = \sum_n \varphi_n S_{np}, \qquad \ldots\ldots(2.13)$$

while

$$S_{np}{}^* = (S^{-1})_{pn}, \qquad \ldots\ldots(2.14)$$

so that

$$\sum_n S_{np}{}^* S_{nq} = \delta_{pq}, \qquad \ldots\ldots(2.15)$$

where δ_{pq} is the Kronecker delta function.

From Eqns (2.12) and (2.13) it follows that

$$c_n{}^k = \sum_q S_{nq} d_q{}^k, \qquad \ldots\ldots(2.16)$$

and using Eqn (2.15) we get from this equation

$$d_q^k = \sum_n c_n^k S_{np}^*. \qquad \ldots\ldots(2.17)$$

If we denote the transformed density matrix by a prime, we have instead of Eqn (2.6) the equation

$$\rho_{pq}' = N^{-1} \sum_k d_q^{k*} d_p^k, \qquad \ldots\ldots(2.18)$$

and from Eqns (2.17) and (2.14) we get

$$\rho_{pq}' = N^{-1} \sum_{k,m,n} c_n^{k*} S_{nq} c_m^k S_{mp}^* = \sum_{m,n} (S^{-1})_{pm} \rho_{mn} S_{nq}, \qquad \ldots\ldots(2.19)$$

or, in matrix notation,

$$\hat{\rho}' = S^{-1} \hat{\rho} S, \qquad \ldots\ldots(2.20)$$

which is the normal equation for the transformation of an operator.

We notice, by the way, that the averages $\langle \hat{A} \rangle$ will be unaffected by the transformation since

$$\langle \hat{A} \rangle' = \mathrm{Tr}\, \hat{\rho}' \hat{A}' = \mathrm{Tr}\, S^{-1} \hat{\rho} S S^{-1} \hat{A} S = \mathrm{Tr}\, S^{-1} \hat{\rho} \hat{A} S = \mathrm{Tr}\, S S^{-1} \hat{\rho} \hat{A} = \mathrm{Tr}\, \hat{\rho} \hat{A} = \langle \hat{A} \rangle, \qquad \ldots\ldots(2.21)$$

where we have used the property of the trace

$$\mathrm{Tr}\, \hat{A} \hat{B} \hat{C} = \mathrm{Tr}\, \hat{B} \hat{C} \hat{A}, \qquad \ldots\ldots(2.22)$$

and the fact that $S S^{-1} = 1$.

Let us now assume that we have made a transformation to a representation in which $\hat{\rho}$ is diagonal,

$$\rho_{mn} = \rho_m \delta_{mn}. \qquad \ldots\ldots(2.23)$$

Consider now $\mathrm{Tr}\, \hat{\rho}^2 (= \langle \hat{\rho} \rangle)$. We have

$$\mathrm{Tr}\, \hat{\rho}^2 = \sum_m \rho_m^2 \leqslant (\sum_m \rho_m)^2 = (\mathrm{Tr}\, \hat{\rho})^2 = 1, \qquad \ldots\ldots(2.24)$$

where we have used Eqn (2.9) and the fact that the ρ_m are non-negative. As the trace is invariant under a unitary transformation, we can write Eqn (2.24) in the form

$$\mathrm{Tr}\, \hat{\rho}^2 \leqslant 1, \quad \text{or,} \quad \sum_{n,m} \rho_{mn} \rho_{nm} = \sum_{n,m} |\rho_{mn}|^2 \leqslant 1, \qquad \ldots\ldots(2.25)$$

where we have used Eqn (2.8). Inequality (2.25) imposes a limitation upon all elements of the density matrix, that is, upon both the diagonal and the off-diagonal elements.

If we are dealing with a finite matrix, say of rank M, there are altogether M^2 complex matrix elements ρ_{mn}, that is, $2M^2$ parameters. This number is reduced by a factor 2 because of Eqn (2.8) and reduced by unity because of the normalization condition (2.9) so that there are $M^2 - 1$ independent parameters. From the operational point of view this means that one needs $M^2 - 1$ independently measured quantities to fix the density matrix.

In the preceding section we have discussed the difference between a pure state and a mixture. We now wish to establish the condition that the density matrix corresponds to a pure state. A pure state corresponds to the case where we have the maximum information available about the physical system, that is, where all systems in the ensemble possess the same wave function, ψ_0 say. In that case there is only one averaging process involved in obtaining $\langle \hat{A} \rangle$, namely, the quantum mechanical one. The pure state is characterized by the existence of what Fano

(1957) calls a ' complete ' experiment ; this is an experiment which gives a result predictable with certainty when performed on this state and gives this particular result only for this particular state. The possibility of such an experiment is apparent if we bear in mind that it is possible to find a Hermitian operator, corresponding to a physical observable, which possesses this particular state as a (non-degenerate) eigenstate. This complete experiment can be used as a filter. It should be an experiment designed to measure the observable, of which ψ_0 is an eigenfunction. We may mention a Nicol prism as a possible apparatus which could serve as a filter when we are dealing with density matrices describing the polarization of light (compare § 8).

We shall now prove that the necessary and sufficient condition that $\hat{\rho}$ describe a pure state is

$$\hat{\rho}^2 = \hat{\rho}, \qquad \ldots\ldots(2.26)$$

that is, $\hat{\rho}$ should be *idempotent*. That Eqn (2.26) is a necessary condition follows easily from the fact that Eqn (2.6) now leads to

$$\rho_{mn} = c_n{}^{(0)*} c_m{}^{(0)}, \qquad \ldots\ldots(2.27)$$

as the $c_n{}^k$ are independent of k and are all equal to the expansion coefficients $c_n{}^{(0)}$ of ψ_0 where ψ_0 is the wave function of all systems in the ensemble. As ψ_0 is normalized we get immediately

$$(\rho^2)_{mn} = \sum_l c_m{}^{(0)} c_l{}^{(0)*} c_l{}^{(0)} c_n{}^{(0)*} = c_m{}^{(0)} c_n{}^{(0)*} = \rho_{mn}. \qquad \ldots\ldots(2.28)$$

To prove that Eqn (2.26) is a sufficient condition we consider again a representation for which $\hat{\rho}$ is diagonal. In that case $\hat{\rho}^2$ is also diagonal, and Eqn (2.26) is now equivalent to

$$\rho_n{}^2 = \rho_n \quad \text{for all values of } n, \qquad \ldots\ldots(2.29)$$

or : $$\text{either} \quad \rho_n = 0 \quad \text{or} \quad \rho_n = 1. \qquad \ldots\ldots(2.30)$$

From the normalization condition (2.9) and Eqn (2.30) it follows that one of the ρ_n, say ρ_0, is equal to unity while the other ρ_n vanish :

$$\rho_0 = 1 \; ; \quad \rho_n = 0, \quad n \neq 0. \qquad \ldots\ldots(2.31)$$

From Eqn (2.6) we then get

$$\rho_0 = N^{-1} \textstyle\sum_k |c_0{}^k|^2 = 1, \qquad \ldots\ldots(2.32)$$

and as all $|c_n{}^k|^2$ must be not greater than 1, Eqn (2.32) can only be satisfied by putting

$$|c_0{}^k| = 1, \quad k = 1, \ldots, N, \qquad \ldots\ldots(2.33)$$

or $$\psi^k = \exp(i\alpha_k)\varphi_0, \quad \alpha_k \text{ real}, \qquad \ldots\ldots(2.34)$$

which means that, apart from an irrelevant phase factor $\exp(i\alpha_k)$, all ψ^k correspond to the same wave function.

We shall now derive the equation of motion for the density matrix. We have assumed that the systems in the ensemble are describable by a wave function and we shall assume that the wave function satisfies the Schrödinger equation

$$\hat{H}\psi^k = i\hbar\dot{\psi}^k, \qquad \ldots\ldots(2.35)$$

where, as usual, the dot indicates differentiation with respect to time. The operator $\hat{H}$ is the Hamiltonian of the system. It is convenient to use Eqn (2.1) to obtain the transformed Schrödinger equation

$$i\hbar\dot{c}_n{}^k = \sum_l H_{nl}\, c_l{}^k, \qquad \ldots\ldots(2.36)$$

and from Eqns (2.36) and (2.6) and the fact that $\hat{H}$ is Hermitean it follows straightforwardly that

$$i\hbar\dot{\rho}_{mn} = [\hat{H}, \hat{\rho}]_{-mn}, \quad \text{or} \quad i\hbar\dot{\hat{\rho}} = [\hat{H}, \hat{\rho}]_-, \qquad \ldots\ldots(2.37)$$

where $[\ ,\ ]_-$ indicates the commutator. We shall defer a discussion of Eqn (2.37) until after we have considered the quantum-mechanical approach to the density matrix, but there are one or two remarks we wish to make at this juncture.

We have so far been working in the Schrödinger representation, that is, the operators are assumed to be time-independent and the time dependence is contained in the wave function (compare Eqn (2.35)). In the Heisenberg representation, on the other hand, the time dependence is in the operators. Instead of a wave (or Schrödinger) equation one has equations of motion for the operators. These are of the same form as the equation of motion (2.37) for $\hat{\rho}$ *but with the opposite sign*. If we were to change to the Heisenberg representation we would find that $\hat{\rho}$ would be a constant operator (see, for instance, Hagedorn 1958).

The time dependence of $\langle\hat{A}\rangle$ is, in the Schrödinger picture, invested in $\hat{\rho}$ and in the Heisenberg picture in $\hat{A}$. For $\langle\hat{A}\rangle$ we find from Eqns (2.7) and (2.37), using the fact that $\hat{A}$ is time-independent in the Schrödinger picture

$$i\hbar\dot{\langle\hat{A}\rangle} = i\hbar\,\overline{\mathrm{Tr}\,\hat{\rho}\hat{A}} = i\hbar\,\mathrm{Tr}\,\dot{\hat{\rho}}\hat{A} = \mathrm{Tr}\,[\hat{H}, \hat{\rho}]_-\hat{A} = \mathrm{Tr}\,(\hat{H}\hat{\rho}\hat{A} - \hat{\rho}\hat{H}\hat{A}),$$

or,

$$i\hbar\dot{\langle\hat{A}\rangle} = \langle[\hat{A}, \hat{H}]_-\rangle, \qquad \ldots\ldots(2.38)$$

with the sign corresponding to the Heisenberg picture! Eqn (2.38) is, of course, true in all representations.

We now turn to the quantum-mechanical approach to the density matrix. We consider a system with many degrees of freedom. We shall denote those degrees of freedom in which we are interested collectively by x and the other degrees of freedom collectively by q. The total wave function of the system will be a function of both q and x: $\Psi(q, x)$. Consider now an operator $\hat{A}$ which is a function of the x only. Its average value will be given by the equation

$$\langle\hat{A}\rangle = \iint \Psi^*(q, x)\,\hat{A}\Psi(q, x)\,dq\,dx, \qquad \ldots\ldots(2.39)$$

where $\int \ldots dq$ and $\int \ldots dx$ denote integrations over the whole of q-space and the whole of x-space.

If we define the density matrix $\hat{\rho}$ by the equation

$$\langle x|\hat{\rho}|x'\rangle = \int \Psi^*(q, x')\,\Psi(q, x)\,dq, \qquad \ldots\ldots(2.40)$$

where we have used for the moment Dirac's notation for operators, Eqn (2.39) is equivalent to

$$\langle\hat{A}\rangle = \iint \langle x'|\hat{\rho}|x\rangle\,dx\,\langle x|\hat{A}|x'\rangle\,dx' = \mathrm{Tr}\,\hat{\rho}\hat{A}, \qquad \ldots\ldots(2.41)$$

where we have used the definition of the trace in coordinate representation.

From Eqn (2.40) and the fact that $\Psi(q, x)$ is normalized we get easily

$$\mathrm{Tr}\,\hat{\rho} = \int \langle x|\hat{\rho}|x\rangle\, dx = \int \Psi^*(q, x)\,\Psi(q, x)\, dx = 1. \quad \ldots\ldots(2.42)$$

We see from Eqn (2.41) that as far as the degrees of freedom which are described by x are concerned, $\hat{\rho}$ describes the situation completely. As Eqns (2.41) and (2.42) are the same as Eqns (2.7) and (2.9), respectively, the density matrices defined by Eqns (2.40) and (2.6) must also be the same. We can pursue this point a little further. The degrees of freedom corresponding to the q may be called the ' bath ' in which the system described by the x is embedded. The averaging over the q takes the place of the earlier ensemble average. To know how $\hat{\rho}$ will change in time, we isolate the system from the bath and consider the Hamiltonian $\hat{H}$ which describes the behaviour of the isolated system. Let $\psi_n(x, t)$ be the complete orthonormal set of time-dependent eigenfunctions of $\hat{H}$. We can expand $\langle x|\hat{\rho}|x'\rangle$ in terms of the ψ_n and write

$$\langle x|\hat{\rho}|x'\rangle = \textstyle\sum_{m,n} a_{mn}\,\psi_n^*(x', t)\,\psi_m(x, t). \quad \ldots\ldots(2.43)$$

We then find for $\hat{\rho}$ the following equation of motion.

$$\begin{aligned} i\hbar\langle x|\dot{\hat{\rho}}|x'\rangle &= \textstyle\sum_{m,n} a_{mn}\, i\hbar[\psi_n^*(x', t)\,\dot{\psi}_m(x, t) + \dot{\psi}_n^*(x', t)\,\psi_m(x, t)] \\ &= \textstyle\sum_{m,n} a_{mn} \int \psi_n^*(x', t)\,\langle x|\hat{H}|x''\rangle\, dx''\,\psi_m(x'', t) \\ &\qquad - \psi_m(x, t)\,\psi_n^*(x'', t)\, dx''\langle x''|\hat{H}^*|x'\rangle \\ &= \int [\langle x|\hat{H}|x''\rangle\, dx''\langle x''|\rho|x'\rangle - \langle x|\rho|x''\rangle\, dx''\langle x''|\hat{H}|x'\rangle] \\ &= \langle x|\hat{H}\hat{\rho} - \hat{\rho}\hat{H}|x'\rangle, \quad \ldots\ldots(2.44) \end{aligned}$$

in accordance with Eqn (2.37).

If we can write

$$\Psi(q, x) = \Phi(q)\,\chi(x), \quad \ldots\ldots(2.45)$$

we are dealing with a pure case. In that case we have

$$a_{mn} = a_m a_n^*, \quad \ldots\ldots(2.46)$$

where the a_m are the coefficients in the expansion of $\chi(x)$ in terms of the $\psi_m(x, t)$. From Eqn (2.40) it follows that

$$\langle x|\hat{\rho}|x'\rangle = \chi^*(x')\,\chi(x), \quad \ldots\ldots(2.47)$$

and we get

$$\begin{aligned} \langle x|\hat{\rho}^2|x'\rangle &= \int \langle x|\hat{\rho}|x''\rangle\, dx''\,\langle x''|\hat{\rho}|x'\rangle \\ &= \int \chi^*(x'')\,\chi(x)\, dx''\,\chi^*(x')\,\chi(x'') = \chi^*(x')\,\chi(x) = \langle x|\hat{\rho}|x'\rangle, \quad \ldots\ldots(2.48) \end{aligned}$$

that is, the condition (2.26) for a pure case is satisfied.

Instead of using the $\psi_n(x, t)$ to expand $\langle x'|\hat{\rho}|x\rangle$ we could have used the complete orthonormal set $\varphi_n(x)$ which we used in Eqn (2.1). In that case we would have got

$$\langle x|\hat{\rho}|x'\rangle = \textstyle\sum_{m,n} c_{mn}\,\varphi_n^*(x')\,\varphi_m(x), \quad \ldots\ldots(2.49)$$

where the c_{mn} possess the same properties as the ρ_{mn} at the beginning of this section. Indeed, one can prove from Eqns (2.41), (2.42) and (2.44) that the c_{mn} satisfy Eqns (2.7), (2.9) and (2.37) where all matrix elements are again in the φ_n-representation. It is known from quantum mechanical transformation theory that the $\varphi_n(x)$ can be used as a transformation ' matrix ' to change over from the discrete n-representation (which we previously called the φ_n-representation) to the continuous x- or coordinate-representation. We see thus that the density matrix introduced in the quantum mechanical method is just the coordinate-representation of the density matrix introduced in the statistical approach.

To conclude this section we shall discuss briefly the case of compound systems (we shall return to this subject in § 10). Let our system consist of two sub-systems A and B, and let superscripts AB, A and B denote quantities referring respectively to the complete system and to the two sub-systems. The density matrix ρ^{AB} will now have two sets of indices, one (m, m') referring to A and the other (n, n') referring to B. It is important to define averages over the sub-systems. We have

$$\langle \hat{A} \rangle_A = \mathrm{Tr}_{AB}(\hat{A}\hat{1}^B\hat{\rho}^{AB}) = \Sigma_{mm'} A_{mm'} \Sigma_{nn'} \delta_{nn'} \rho^{AB}_{m'n'mn} = \mathrm{Tr}_A \hat{A}\hat{\rho}^A \qquad \ldots\ldots(2.50)$$

where the density matrix $\hat{\rho}^A$ for the sub-system A is defined by

$$\hat{\rho}^A = \mathrm{Tr}_B \hat{\rho}^{AB} \quad \text{or} \quad \rho^A_{mm'} = \Sigma_n \rho^{AB}_{mnm'n}, \qquad \ldots\ldots(2.51)$$

and where $\hat{A}$ is an operator acting upon the sub-system A only.

If the two sub-systems are completely independent, we must have for any two operators $\hat{A}$ (acting upon A) and $\hat{B}$ (acting upon B)

$$\langle \hat{A}\hat{B} \rangle_{AB} = \langle \hat{A} \rangle_A \langle \hat{B} \rangle_B, \qquad \ldots\ldots(2.52)$$

and the density matrix must satisfy the condition

$$\hat{\rho}^{AB} = \hat{\rho}^A \times \hat{\rho}^B \quad \text{or} \quad \rho^{AB}_{mnm'n'} = \rho^A_{mm'} \rho^B_{nn'}, \qquad \ldots\ldots(2.53)$$

where $\times$ indicates a direct product.

§ 3. Green Function Techniques

In this section we shall briefly consider the definition and properties of the temperature-dependent Green functions which are playing such an important role in recent developments of quantum statistical mechanics. We cannot enter into a detailed discussion here, and refer to the steadily growing literature, especially to the papers by Matsubara (1955 b), Martin and Schwinger (1959) and Zubarev (1960). Other relevant papers are those by Koppe (1951), Salam (1953), Kinoshita and Nambu (1954), Bonch-Bruevich (1955, 1956 a, b, c, 1957, 1958 a, b, 1959 a, b, c, d), Van Hove (1955, 1957), Watson (1956), Ezawa, Tomozawa and Umezawa (1957), Klein and Zemach (1957), Kubo (1957 a), Migdal (1957), Belyaev (1958 a, b), Galitskiĭ (1958), Galitskiĭ and Migdal (1958), Gor'kov (1958), Klein and Prange (1958), Kobelev (1958 a, b, c, d, e, 1959), Kraichnan (1958 a, b), Landau (1958), Montroll and Ward (1958, 1959), Prange and Klein (1958), Abrikosov, Gor'kov and Dzyaloshinskiĭ (1959), Bogolyubov and Tyablikov (1959), Bonch-Bruevich and Glasko (1959), Chen' Chun'-syan' (1959), Falk (1959 a, b), Fradkin (1959 a, b, c, d, e), Fujita (1959), Hugenholtz and Pines (1959), Kanazawa

and Watabe (1959), Klinger (1959 a), Kogan (1959), Pitaevskiĭ (1959), Vedenov (1959), Vedenov and Larkin (1959), Balian and De Dominicis (1960 a), Bonch-Bruevich and Kogan (1960), Bonch-Bruevich and Mironov (1960), Dzyub (1960), Fujita and Hirota (1960) and Sawicki (1960).

We mentioned in the introduction that the use of Green functions came about quite naturally on two accounts. First of all, the resemblance between the main statistical problem—that of evaluating the thermodynamic properties of a system of interacting particles— and the main field-theoretical problem—that of considering interacting fields—when both problems are stated in terms of annihilation and creation operators leads naturally to the use of Green functions in quantum statistics, as Matsubara (1955 b) was the first to note. Secondly, the close relation between the correlation functions of kinetic theory and the two- or more-particle Green functions (*vide infra*) makes their occurrence also a natural one. In this section we shall briefly introduce the Green functions and briefly consider some of their properties which will be of interest for the discussions in later sections.

There are many ways of defining Green functions, and it is perhaps expedient to dwell briefly on the connection between different Green functions. There are two kinds of Green functions in which we shall be interested, namely, the generalized double-time (causal) Green functions discussed by Zubarev (1960 ; it was pointed out by Zubarev that it is often more convenient to use the advanced or the retarded Green functions, because of their analytical properties, but this distinction is too subtle for our present discussion), and the propagators of field theory. The generalized Green function of two operators $\hat{A}$ and $\hat{B}$ is defined by the equation

$$G_{\hat{A},\hat{B}}(t,t') = (-i/\hbar)\langle T\hat{A}(t)\,\hat{B}(t')\rangle, \qquad \ldots\ldots(3.1)$$

where $\hat{A}$ and $\hat{B}$ are time-dependent operators, and where T indicates Wick's chronological or the T-product of the two operators. One can write this T-product in the form

$$T\hat{A}(t)\,\hat{B}(t') = \theta(t-t')\,\hat{A}(t)\,\hat{B}(t') + \eta\theta(t'-t)\,\hat{B}(t')\,\hat{A}(t), \quad \ldots\ldots(3.2)$$

where θ is the step function

$$\theta(t) = 1, \quad t > 0\,; \qquad \theta(t) = 0, \quad t < 0, \qquad \ldots\ldots(3.3)$$

and where $\eta = +1$ or -1 according to whether $\hat{A}$ and $\hat{B}$ are Bose- or Fermi-operators. The generalization of the usual Green function of field theory consists of two parts. Firstly, we have not yet made any restrictions on the operators $\hat{A}$ and $\hat{B}$, while in field theory $\hat{A}$ and $\hat{B}$ are normally the second-quantized wave functions, that is, the wave functions which are themselves operators by virtue of their expansion in creation and annihilation operators. Secondly, the average in Eqn (3.1) is over a statistical ensemble, in accordance with our definition of the $\langle\ \rangle$ average. It is usual to choose a canonical or grand canonical ensemble for this average, that is, an equilibrium ensemble, but in our definition this restriction has not yet been introduced. In field theory, the average is normally over a pure state, namely, the ground state. There is also one other difference, namely, that the Green functions which we have introduced depend on two time arguments only, while in field theory there occur multiple-time Green functions.

The propagator Green function occurs in field theory when one discusses diagram techniques (see, for instance, Bogolyubov and Shirkov 1959 or Hamilton 1959) and it is then shown that, indeed, it coincides with the usual definition of a Green function in the theory of differential equations. This definition is the following one (for instance, Morse and Feshbach 1953). If we wish to obtain the field due to a source distribution, we calculate the effect of each part of the source distribution and add them up. If $G(\mathbf{x}, t\,;\, \mathbf{x}', t')$ is the field at $\mathbf{x}$ at time t caused by a unit point source at $\mathbf{x}'$ at time t', $G(\mathbf{x}, t\,;\, \mathbf{x}', t')$ is called the Green function of the problem considered. If the field equation is given by the equation

$$\Omega\psi = f, \qquad \ldots\ldots(3.4)$$

where ψ is the field function (for instance, the components of the electromagnetic four-potential), Ω a differential operator (in the case of the electromagnetic field the d'Alembertian), and f the source density (in the electromagnetic case proportional to the current–charge four-vector components), the Green function will be a solution of the equation

$$G(\mathbf{x}, t\,;\, \mathbf{x}', t') = \alpha\delta(\mathbf{x}-\mathbf{x}')\,\delta(t-t'), \qquad \ldots\ldots(3.5)$$

where $\delta(\mathbf{x})$ and $\delta(t)$ are the three- and one-dimensional Dirac delta-functions. The constant α depends on the exact form of the right-hand side of Eqn (3.4).

We shall try to make it plausible how these Green functions play the role of propagators by considering the Fokker–Planck equation. This is the diffusion equation for Brownian motion. The ordinary diffusion equation has a field operator Ω given by the equation

$$\Omega = \nabla^2 - a^2\,\partial/\partial t, \qquad \ldots\ldots(3.6)$$

where a is a constant involving the diffusion coefficient. The corresponding Green function equation

$$(\partial G/\partial t) - a^{-2}\nabla^2 G = \delta(\mathbf{x}-\mathbf{x}')\,\delta(t-t'), \qquad \ldots\ldots(3.7)$$

is the Fokker–Planck equation for the probability that a particle at time t' at $\mathbf{x}'$ will at time t be at $\mathbf{x}$ through Brownian motion. The Green function thus describes the propagation of the Brownian particle through the medium.

For our further discussion it is instructive to see how the Green function (3.1) for the special choice of $\hat{A} = \hat{\psi}$, $\hat{B} = \hat{\psi}^\dagger$ satisfies Eqn (3.5). The $\hat{\psi}$ and $\hat{\psi}^\dagger$ are the wave functions in second quantization, and the $\dagger$ indicates, as usual, the Hermitian conjugate, which means that if $\hat{\psi}$ contains creation operators, $\hat{\psi}^\dagger$ will contain annihilation operators and vice versa. The $\hat{\psi}$ and $\hat{\psi}^\dagger$ are thus operators and they satisfy the commutation relations

$$\hat{\psi}(\mathbf{x}, t)\hat{\psi}^\dagger(\mathbf{x}', t') - \eta\hat{\psi}^\dagger(\mathbf{x}', t')\hat{\psi}(\mathbf{x}, t) = \delta(\mathbf{x}-\mathbf{x}')\,\delta(t-t'). \qquad \ldots\ldots(3.8)$$

If $\hat{\psi}$ describes a free field, that is, if there are no interactions, it satisfies the equation of motion

$$\hat{\Omega}\hat{\psi} \equiv i\hbar(\partial\hat{\psi}/\partial t) + (\hbar^2/2m)\nabla^2\hat{\psi} = 0. \qquad \ldots\ldots(3.9)$$

From Eqns (3.1), (3.2), (3.3), (3.8) and (3.9) it follows in a straightforward manner that

$$\hat{\Omega}G_{\hat{\psi},\hat{\psi}^\dagger} = \delta(\mathbf{x}-\mathbf{x}')\,\delta(t-t'), \qquad \ldots\ldots(3.10)$$

where we have used Eqn (2.9) and the fact that

$$-\partial\theta(t'-t)/\partial t = \partial\theta(t-t')/\partial t = \delta(t-t'). \qquad \ldots\ldots(3.11)$$

It is of interest for statistical applications of Green functions to consider also the case of interacting particles. Instead of Eqn (3.9) we then have the equation of motion

$$i\hbar\frac{\partial\hat{\psi}(\mathbf{x},t)}{\partial t}+\frac{\hbar^2}{2m}\nabla^2\hat{\psi}-\int d^3\mathbf{x}'\,v(\mathbf{x},\mathbf{x}')\hat{\psi}^\dagger(\mathbf{x}',t)\hat{\psi}(\mathbf{x}',t)\psi(\mathbf{x},t)=0, \qquad \ldots\ldots(3.12)$$

where we have assumed two-body interactions, and where $v(\mathbf{x},\mathbf{x}')$ is the potential energy corresponding to these interactions. If we now proceed as before we find the following equation of motion for the Green function $G_{\hat{\psi},\hat{\psi}^\dagger}$,

$$i\hbar\frac{\partial G_{\hat{\psi},\hat{\psi}^\dagger}}{\partial t}+\frac{\hbar^2}{2m}\nabla^2 G_{\hat{\psi},\hat{\psi}^\dagger}-\frac{i\eta}{\hbar}\int d^3\mathbf{x}''\,v(\mathbf{x},\mathbf{x}'')\,G_{\hat{\psi},\hat{\psi}_*,\hat{\psi},\hat{\psi}_*^\dagger}=\delta(\mathbf{x}-\mathbf{x}')\,\delta(t-t'). \qquad \ldots\ldots(3.13)$$

We note that the equation of motion for the one-particle Green function $G_{\hat{\psi},\hat{\psi}^\dagger}$ involves the two-particle Green function $G_{\hat{\psi},\hat{\psi}_*,\hat{\psi},\hat{\psi}_*^\dagger}$. In turn the equation of motion for the two-particle Green function will involve the three-particle Green function, and so on. This situation is similar to the situation in the theory of transport processes (compare § 7) and in the theory of liquids where we encounter similar chains of equations for the correlation functions.

Let us briefly consider this connection between the Green functions and the correlation functions $F_{\hat{A},\hat{B}}$ which are defined by the equation

$$F_{\hat{A},\hat{B}}(t,t')=\langle\hat{A}(t'),\hat{B}(t)\rangle. \qquad \ldots\ldots(3.14)$$

Note that, in general, $F_{\hat{A},\hat{B}}\neq F_{\hat{B},\hat{A}}$, just as $G_{\hat{A},\hat{B}}\neq G_{\hat{B},\hat{A}}$, and also that $F_{\hat{A},\hat{B}}$ is defined for $t=t'$, which is not the case for $G_{\hat{A},\hat{B}}$. The close connection between the $F_{\hat{A},\hat{B}}$ and the $G_{\hat{A},\hat{B}}$ is apparent from Eqns (3.14), (3.1) and (3.2). The correlation functions (and thus the Green functions) describe the microscopic properties of physical systems. As statistical mechanics is interested in deriving the macroscopic properties from a knowledge of the microscopic behaviour, we see still another reason why Green function techniques have become important. We shall not discuss one of the most important applications of Green functions—an application for which it is essential to use the time-dependent Green functions—namely, the derivation of the energy spectrum of the elementary excitations, that is, of the quasi-particles which originate from the many-body interactions in the system.

The Green functions will occur in the next section as propagators, and also in §§ 6 and 7 as correlation functions.

§ 4. The Description of Statistical Equilibrium by the Density Matrix

In this section we shall first of all discuss the form of the density matrix which corresponds to thermodynamic equilibrium (see Tolman 1938, Husimi 1940, ter Haar 1954, 1955). We also discuss the relation between this density matrix and the various thermodynamic functions of the system under consideration. We shall then consider some of the methods developed recently for the evaluation of

the grand canonical or canonical partition function. Among these methods we may mention the Feynman path method (Feynman 1953, Yaglom 1956, Husimi, Kitano and Nishiyama 1958, Brush 1961, Siegert 1960), Feynman graph methods (Matsubara 1955 b, Riesenfeld and Watson 1956, Thouless 1957, 1959, Bloch and De Dominicis 1958, 1959 a, b, Montroll and Ward 1958, Brout 1959, Glassgold, Heckrotte and Watson 1959, Lee and Yang 1959, Balian and De Dominicis 1960 a, Levine 1960, Stillinger and Kirkwood 1960), and second quantization methods (Blatt and Matsubara 1958, Fujita 1959, Hubbard 1959, Balian and De Dominicis 1960 b, c, Gaudin 1960 b). For other developments, especially perturbation theory methods, we can refer to papers by Goldberger and Adams (1952), Kubo (1952), Chester (1954), Takabayasi (1954), Zubarev (1954), Butler and Friedman (1955), Friedman and Butler (1955), Kaschluhn (1955 a), Kotani (1955), Nakajima (1955), Watson (1956), Mazur and Oppenheim (1957), Oppenheim and Mazur (1957), Oppenheim and Ross (1957), Schafroth, Butler and Blatt (1957), Yvon (1957, 1960 a, b), Klein and Prange (1958), Pavlikovskiĭ and Shchuruvna (1958, 1959), Siegert and Teramoto (1958), Lundquist (1959), Balian, Bloch and De Dominicis (1960), Fujita and Hirota (1960) and Gaudin (1960 a), and also to many of the papers quoted in the previous section. We may remark here that a number of authors have applied graph methods to classical partition functions with great success (for instance, Salpeter 1958, van Leeuwen, Groeneveld and de Boer 1959).

If the system is in a state of thermodynamic equilibrium, the density matrix must be time-independent, and we see from Eqn (2.37) that it must satisfy the equation

$$[\hat{H}, \hat{\rho}]_- = 0, \qquad \ldots\ldots(4.1)$$

which is most easily attained by letting $\hat{\rho}$ be a function of $\hat{H}$. We have mentioned before that one way of looking at an ensemble is by considering one of the systems in the ensemble to be embedded in a bath formed by the other systems. Just as the distribution function of a molecule in a gas will be given by the Maxwell–Boltzmann distribution which is produced by the interaction of the molecule with the other molecules in the gas, so the density matrix of an ensemble in thermodynamic equilibrium will be given by the canonical or grand canonical expression

$$\hat{\rho} = \exp[-q + \nu\hat{N} - \beta\hat{H}]. \qquad \ldots\ldots(4.2)$$

In Eqn (4.2) q, ν and β are c-numbers, and $\hat{N}$ is the number operator. We have given here the formula for the grand canonical ensemble, that is, an ensemble where the constituent systems may have different numbers of particles. One can give a more detailed proof that Eqn (4.2) corresponds to an equilibrium ensemble (for instance, ter Haar 1955). We must also refer to the literature for a proof of the statement that β is related to the absolute temperature T by the equation

$$\beta^{-1} = kT, \qquad \ldots\ldots(4.3)$$

where k is Boltzmann's constant, and that $\nu = \beta\mu$, where μ is the chemical or thermal potential. The q-potential, finally, is determined from Eqn (2.9), or

$$\exp q = \mathrm{Tr}\exp(\nu\hat{N} - \beta\hat{H}). \qquad \ldots\ldots(4.4)$$

We also note, without proof, that the grand potential q determines all physical quantities for a grand canonical ensemble just as the free energy does for a canonical ensemble. The quantity Z defined by the equation

$$Z = \exp q, \qquad \ldots\ldots(4.5)$$

is called the *grand partition function* and is the quantity which is usually discussed in the papers quoted at the beginning of this section.

From Eqn (4.2) it follows that $\hat{\rho}$ satisfies the so-called Bloch equation (Bloch 1932)

$$\partial\hat{\rho}/\partial\beta = -\hat{H}\hat{\rho}. \qquad \ldots\ldots(4.6)$$

The fact that this equation is formally the same as the Schrödinger equation, with β replacing $it/\hbar$, has been the basis for recent developments. Feynman's path integral method (Feynman 1948) was applied by Feynman himself to liquid helium (Feynman 1953) and has recently been discussed by Brush (1961). Watson (1956) has used this analogy to apply scattering theory for an evaluation of the second virial coefficient. Let $\hat{R}(E)$ be the Laplace transform of $e^{-\beta H}(= \hat{p}(\beta))$:

$$\hat{R}(E) = -\int_0^\infty e^{\beta E}\,\hat{p}(\beta)\,d\beta, \qquad \ldots\ldots(4.7)$$

and let the Hamiltonian be split into the kinetic energy $\hat{T}$ and the potential energy $\hat{V}$,

$$\hat{H} = \hat{T} + \hat{V}. \qquad \ldots\ldots(4.8)$$

We then have from Eqn (4.6), which is also satisfied by $\hat{p}(\beta)$,

$$(E - \hat{T})\,\hat{R} = 1 + \hat{V}\hat{R}, \qquad \ldots\ldots(4.9)$$

where we have used the fact that $\hat{p}(0) = 1$.

From Eqn (4.9) we get the equation

$$\hat{R} = (E - \hat{T})^{-1} + (E - \hat{T})^{-1}\,\hat{V}\hat{R}. \qquad \ldots\ldots(4.10)$$

This equation reduced to the equation of stationary state scattering

$$\hat{\Omega} = 1 + (E - \hat{T})^{-1}\,\hat{V}\hat{\Omega}, \qquad \ldots\ldots(4.11)$$

if we write

$$\hat{\Omega}(E) = \hat{R}(E)(E - \hat{T}). \qquad \ldots\ldots(4.12)$$

The equation for the density matrix is thus reduced to the scattering equation and quantum-mechanical scattering theory (for instance, Brueckner theory : see Bethe 1956) can be applied. We have not got the space to discuss the many theories which have, indeed, used Brueckner theory to evaluate Z.

We note from Eqn (4.10) that $\hat{R}$ is the Green function for the operator $E - \hat{H}$ (see Koppe 1951).

We shall now briefly discuss the Feynman diagram methods, but once again we cannot go into a detailed discussion. We refer especially to Gaudin's paper (1960 b) which is based upon the earlier work of Matsubara (1955 b) and Bloch and De Dominicis (1958); he seems to have given the clearest exposé of this rather complicated method. For many details we must refer to textbooks on quantum field theory (for instance, Bogolyubov and Shirkov 1959, Hamilton 1959).

We shall consider a system of bosons which interact through two-body forces. The Hamiltonian is then

$$\hat{H} = \hat{H}_0 + \hat{V}, \qquad \ldots\ldots(4.13)$$

where $\hat{H}_0$ corresponds to non-interacting particles and $\hat{V}$ to the interaction. If we describe the system in terms of creation and annihilation operators, $\hat{a}_k^\dagger$ and $\hat{a}_k$, where $\hat{a}_k^\dagger$ creates a particle of energy ϵ_k and $\hat{a}_k$ annihilates such a particle, we have

$$\hat{H}_0 = \Sigma_k \epsilon_k \hat{a}_k^\dagger \hat{a}_k, \qquad \ldots\ldots(4.14)$$

and

$$V = \tfrac{1}{2}\Sigma_{r,s,m,n} \langle rs|v|mn\rangle \hat{a}_r^\dagger \hat{a}_s^\dagger \hat{a}_m \hat{a}_n, \qquad \ldots\ldots(4.15)$$

where we have once again used Dirac's notation for the matrix elements of the two-body interaction potential $v(\mathbf{x}, \mathbf{x}')$. As we are dealing with bosons the $\hat{a}_k$ and $\hat{a}_k^\dagger$ satisfy the commutation relations (compare Eqn (3.7))

$$[\hat{a}_k, \hat{a}_{k'}]_- = [\hat{a}_k^\dagger, \hat{a}_{k'}^\dagger]_- = 0, \quad [\hat{a}_k, \hat{a}_{k'}^\dagger]_- = \delta_{kk'}. \qquad \ldots\ldots(4.16)$$

We now write

$$\hat{\rho}(\beta) = \hat{\rho}_0(\beta)\hat{S}(\beta), \qquad \ldots\ldots(4.17)$$

where again

$$\hat{\rho}(\beta) = e^{-\beta\hat{H}}, \qquad \ldots\ldots(4.18)$$

while

$$\hat{\rho}_0(\beta) = \exp(-\beta\hat{H}_0). \qquad \ldots\ldots(4.19)$$

As $\hat{\rho}(\beta)$ satisfies the Bloch equation, we find that $\hat{S}(\beta)$ must satisfy the equation

$$-\partial\hat{S}/\partial\beta = \hat{V}(\beta)\hat{S}(\beta), \qquad \ldots\ldots(4.20)$$

where

$$\hat{V}(\beta) = \exp(\beta\hat{H}_0)\,\hat{V}\exp(-\beta\hat{H}_0). \qquad \ldots\ldots(4.21)$$

It also satisfies the boundary condition

$$\hat{S}(0) = 1, \qquad \ldots\ldots(4.22)$$

and we find the following solution of Eqn (4.20) for $\hat{S}$,

$$\hat{S}(\beta) = \sum_{n=0}^{\infty}(-1)^n \int_0^\beta d\beta_1 \int_0^{\beta_1} d\beta_2 \ldots \int_0^{\beta_{n-1}} d\beta_n\, \hat{V}(\beta_1)\ldots\hat{V}(\beta_n). \qquad \ldots\ldots(4.23)$$

For the grand partition function Z we have

$$\begin{aligned} Z &= \mathrm{Tr}\,[\exp(\nu\hat{N} - \beta\hat{H})] \\ &= \mathrm{Tr}\,[\exp(\nu\hat{N} - \beta\hat{H}_0).\hat{S}(\beta)] \\ &= \mathrm{Tr}\,[\exp(\nu\hat{N} - \beta\hat{H}_0)]\,\langle\hat{S}(\beta)\rangle_0, \end{aligned} \qquad \ldots\ldots(4.24)$$

where $\langle\ldots\rangle_0$ indicates an average over an ensemble with $\hat{H} = \hat{H}_0$, that is,

$$\begin{aligned} \langle\hat{S}(\beta)\rangle_0 &= \mathrm{Tr}\,[\exp(-q_0 + \nu\hat{N} - \beta\hat{H}_0).\hat{S}(\beta)] \\ &= \{\mathrm{Tr}\,[\exp(\nu\hat{N} - \beta\hat{H}_0).\hat{S}(\beta)]\}\{\mathrm{Tr}\,[\exp(\nu\hat{N} - \beta\hat{H}_0)]\}^{-1}. \end{aligned} \qquad \ldots\ldots(4.25)$$

We see that we have obtained here an expression of the partition function in a power series in $\hat{V}$, as $\hat{S}$ itself is expressed as such a power series. Using the normal field theoretical techniques we can write Eqn (4.24) in the form

$$Z = Z_0 \sum_{n=0}(-1)^n (n!)^{-1} \int_0^\beta d\beta_1 \int_0^\beta d\beta_2 \ldots \int_0^\beta d\beta_n \langle T\hat{V}(\beta_1)\ldots\hat{V}(\beta_n)\rangle_0, \qquad \ldots\ldots(4.26)$$

where T indicates again the chronological product and where

$$Z_0 = \mathrm{Tr}\,[\exp(\nu\hat{N} - \beta\hat{H}_0)]. \qquad \ldots\ldots(4.27)$$

Using Eqns (4.26), (4.15) and (4.21) we have finally

$$Z = Z_0 \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \cdot \tfrac{1}{4} \sum_{r_1 s_1 m_1 n_1} \langle r_1 s_1 | v | m_1 n_1 \rangle \langle r_2 s_2 | v | m_2 n_2 \rangle \ldots \int_0^\beta d\beta_1 \int_0^\beta d\beta_2 \ldots$$
$$\times \int_0^\beta d\beta_n \langle T\hat{a}_{r_1}^\dagger(\beta_1)\, \hat{a}_{s_1}^\dagger(\beta_1)\, \hat{a}_{n_1}(\beta_1)\, \hat{a}_{m_1}(\beta_1)\, \hat{a}_{r_2}^\dagger(\beta_2)\, \hat{a}_{s_2}^\dagger(\beta)_2 \ldots$$
$$\times \hat{a}_{n_n}^\dagger(\beta_n)\, \hat{a}_{n_n}(\beta_n)\, \hat{a}_{m_n}(\beta_n) \rangle_0. \qquad \ldots\ldots(4.28)$$

It is this expansion for Z which is evaluated by diagram techniques. It must be emphasized here that using diagram techniques does not solve the problem; they only serve as an aide-mémoire so that it is more difficult to forget terms, and also it is often possible to combine a large number of terms and sum those. If one can convince oneself that those terms are, indeed, the only important ones, one hopes to have obtained a reasonable approximation to the exact partition function.

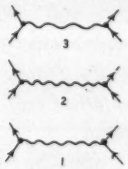


Fig. 1.

To conclude this section we shall show how one can attribute diagrams to the various terms in Eqn (4.28). One first of all uses Wick's theorem (1950; see also Gaudin 1960 a) to sort out the various terms of nth order according to whether the indices of the creation and annihilation operators are different or the same. One is finally left with a product of 'contractions' of the form $\langle T\hat{a}_k^\dagger(\beta)\, \hat{a}_{k'}(\beta') \rangle_0$ and these are equal to the Green functions (compare Eqn (3.1)),

$$\langle T\hat{a}_k^\dagger(\beta)\, \hat{a}_{k'}(\beta') \rangle_0 = i\hbar\, G_{\hat{a}_k^\dagger, \hat{a}_{k'}}(\beta, \beta'). \qquad \ldots\ldots(4.29)$$

To find all diagrams corresponding to the nth order term one proceeds as follows (compare Goldstone 1957 and Hugenholtz 1957; for the sake of simplicity we shall take $n = 3$). Draw 3 horizontal wavy lines (see Fig. 1) numbered 1, 2 and 3, corresponding to the interactions $\hat{V}(\beta_1)$, $\hat{V}(\beta_2)$ and $\hat{V}(\beta_3)$. On the left one draws two lines, one joining from below corresponding to $\hat{a}_m$ and one joining from above corresponding to $\hat{a}_r^\dagger$; similarly on the right two lines corresponding to $\hat{a}_n$ and $\hat{a}_s^\dagger$. This corresponds to a scattering where two particles in states ϵ_m and ϵ_n are scattered into the states ϵ_r and ϵ_s. We shall forget for the sake of simplicity the fact that there is a symmetry between m and n, and between r and s (for a discussion of this point see, for instance, Gaudin 1960 b). The only terms which will contribute to the sum are those where there are no loose ends, and neglecting the symmetry

effects, just mentioned, we see that only diagrams of the kind depicted in Fig. 2 contribute to the term of third order. We do not wish to pursue this discussion any further, but refer to the literature.

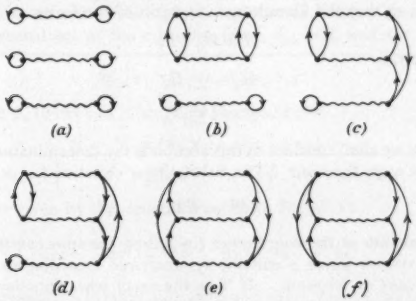


Fig. 2.

§ 5. Density Matrix Techniques applied to Atoms, Molecules and Nuclei

Until now we have been concerned with the properties of the density matrix, rather than with the density matrix as a means of tackling physical problems. This will be the subject of §§ 5 to 9. We start with a discussion of many-body systems in this section and the next. In the present section we shall discuss the application of density matrix techniques to atoms, molecules and nuclei, that is, their application in the Hartree–Fock and Thomas–Fermi–Dirac theories. In that case we are dealing with temperature independent density matrices, or, perhaps more exactly, with density matrices applied to systems at absolute zero, where the systems are in their ground state. Quantum chemists have recently become interested in the use of density matrices, especially the so-called reduced density matrices, and as a result the literature is now growing fast. We may refer here to papers by Dirac (1930 b, 1931), Husimi (1940), Corson (1951), Kinoshita and Nambu (1954), Löwdin (1954, 1955 a, b, c, 1959, 1960), Kaschluhn (1955 b), Macke (1955 a, b), Matsubara (1955 b), Mayer (1955), McWeeny (1955, 1956 a, b, 1957, 1959 a, b, 1960), Blatt (1956), Gombás (1956), Chirgwin (1957), Golden (1957 a, b, 1960 a, b), Koppe (1957), Mizuno and Izuyama (1957), Tredgold (1957), Ayres (1958), Blatt and Matsubara (1958), Husimi, Kitano and Nishiyama (1958), Kobelev (1958 a, e), March and Young (1958), McConnell (1958), Saroléa and Koppe (1958), Bopp (1959), Ehrenreich and Cohen (1959), Fradkin (1959 c), Goldstone and Gottfried (1959), Hubbard (1959), Martin and Schwinger (1959), Naqvi (1959), Peacock and McWeeny (1959), Shull (1959), Alfred (1960), ter Haar (1960), Hall (1960), McWeeny and Mizuno (1961), McWeeny and Ohno (1960), Penrose (1960), Thouless (1961) and Young and March (1960). In writing this section we have also used repeatedly unpublished notes of lectures given by

L. Rosenfeld at Copenhagen in 1958–59. For a general discussion of the Hartree–Fock and the Thomas–Fermi theories we refer to the literature (for instance, Gombás 1956, Hund 1956, Löwdin 1956, March 1957, ter Haar 1958).

In the present section we consider again systems of particles interacting through two-body forces so that the Hamiltonian is again of the form

$$\hat{H} = \hat{H}_0 + \hat{V}, \qquad \ldots\ldots(5.1)$$

where we now write

$$\hat{H}_0 = \Sigma_i \hat{H}_i, \qquad \ldots\ldots(5.2)$$

$$\hat{V} = \tfrac{1}{2}\Sigma_{ij} \hat{V}_{ij}. \qquad \ldots\ldots(5.3)$$

The problem we shall consider in this section is the determination of the ground state energy and wave function. The Schrödinger equation is

$$\hat{H}\Psi = E\Psi, \qquad \ldots\ldots(5.4)$$

where Ψ is a function of the coordinates (including the spin coordinates) of all N particles in the system which is suitably symmetrized according to whether we are dealing with bosons or fermions. If Ψ is the exact wave function, the energy is given by the equation

$$E = \int \Psi^* \hat{H} \Psi \, d\tau_N. \qquad \ldots\ldots(5.5)$$

Using Eqns (5.1) to (5.3) and the fact that Ψ is completely symmetric or antisymmetric in all particles, we can write Eqn (5.5) in the form

$$E = \tfrac{1}{2}N \int \Psi^* [\hat{H}_1 + \hat{H}_2 + (N-1)\hat{V}_{12}] \Psi \, d\tau_N. \qquad \ldots\ldots(5.6)$$

We can now introduce the one- and two-particle density matrices through the equations (compare Eqn (2.40) and bear in mind that both q and x stood for sets of coordinates)

$$\langle \mathbf{x}_1 | \hat{\rho}^{(1)} | \mathbf{x}_1' \rangle = \int \Psi^*(\mathbf{x}_1', \mathbf{x}_2, \ldots, \mathbf{x}_N) \Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) \, d^3\mathbf{x}_2 \ldots d^3\mathbf{x}_N, \qquad \ldots\ldots(5.7)$$

$$\langle \mathbf{x}_1, \mathbf{x}_2 | \hat{\rho}^{(2)} | \mathbf{x}_1', \mathbf{x}_2' \rangle = \int \Psi^*(\mathbf{x}_1', \mathbf{x}_2', \mathbf{x}_3, \ldots, \mathbf{x}_N) \Psi(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_N) \, d^3\mathbf{x}_3 \ldots d^3\mathbf{x}_N, \qquad \ldots\ldots(5.8)$$

where here and henceforth we shall omit explicit reference to spin coordinates; if necessary $\int \ldots d^3\mathbf{x}$ will be assumed to include summation over spin coordinates and $\mathbf{x}$ will be assumed to stand for both spatial and spin coordinates.

From Eqns (5.6), (5.7) and (5.8) we get

$$E = N \operatorname{Tr} \hat{\rho}^{(1)} \hat{H}_1 + \tfrac{1}{2}N(N-1) \operatorname{Tr} \hat{\rho}^{(2)} \hat{V}_{12}, \qquad \ldots\ldots(5.9)$$

or

$$E = \tfrac{1}{2}N \operatorname{Tr} \hat{\rho}^{(2)} \hat{h}, \qquad \ldots\ldots(5.10)$$

where

$$\hat{h} = \hat{H}_1 + \hat{H}_2 + (N-1)\hat{V}_{12}. \qquad \ldots\ldots(5.11)$$

We see that the energy is completely determined by the two-particle density matrix, provided there are no three- or more-body forces.

So far we have not made any assumptions about the form of Ψ, and all equations are exact. The first application of our equations will be one discussed by Bopp (1959). We note that we have expressed the energy exactly in terms of the two-body energy operator (5.11), and one might ask whether one could reduce the N-particle problem to a two-body problem. Let us assume that we know the complete orthonormal set of the eigenfunctions, ϕ_n, of $\hat{h}$ and the corresponding eigenvalues, ϵ_n,

$$\hat{h}\phi_n(\mathbf{x}_1, \mathbf{x}_2) = \epsilon_n \phi_n(\mathbf{x}_1, \mathbf{x}_2). \qquad \ldots\ldots(5.12)$$

We can use these ϕ_n to expand Ψ and also to expand $\hat{\rho}^{(2)}$,

$$\Psi = \sum_n \phi_n(\mathbf{x}_1, \mathbf{x}_2)\, \chi_n(\mathbf{x}_3, \ldots, \mathbf{x}_N), \qquad \ldots\ldots(5.13)$$

$$\langle \mathbf{x}_1, \mathbf{x}_2 | \hat{\rho}^{(2)} | \mathbf{x}_1', \mathbf{x}_2' \rangle = \sum_{n,m} \phi_n(\mathbf{x}_1, \mathbf{x}_2)\, \phi_m^*(\mathbf{x}_1', \mathbf{x}_2')\, p_{nm}, \qquad \ldots\ldots(5.14)$$

where the p_{mn} are given by the equation (compare Eqn (5.8))

$$p_{nm} = \int \chi_n(\mathbf{x}_3, \ldots, \mathbf{x}_N)\, \chi_m^*(\mathbf{x}_3, \ldots, \mathbf{x}_N)\, d^3\mathbf{x}_3 \ldots d^3\mathbf{x}_N. \qquad \ldots\ldots(5.15)$$

Substituting expression (5.14) into Eqn (5.8) and using Eqn (5.12) we see that only the terms with p_{nn} remain in the expression for E and we find

$$E = \tfrac{1}{2}N \sum_n p_{nn}\, \epsilon_n. \qquad \ldots\ldots(5.16)$$

We have, of course, not yet gained anything. We may know the ϵ_n from solving the two-body problem (5.12) but the p_{nn} can only be found from a knowledge of the N-body wave function (see Eqn (5.15)). From Eqns (5.13) and (5.15) we see that p_{nn} is the probability amplitude of the state ϕ_n in Ψ. One can, however, find a good approximation to E by using an argument which is similar to the one used in explaining the electronic structure of the elements in the periodic system. In the latter case one fills up the lowest N levels in the self-consistent field. Similarly one can argue that one should fill up the lowest $\frac{1}{2}N(N-1)$ pair-states, that is, put

$$\begin{aligned} p_{nn} &= 2[N(N-1)]^{-1}, \quad && n = 1, \ldots, \tfrac{1}{2}N(N-1)\,; \\ p_{nn} &= 0, && n > \tfrac{1}{2}N(N-1), \end{aligned} \qquad \ldots\ldots(5.17)$$

which satisfies the normalization condition

$$\sum_n p_{nn} = 1. \qquad \ldots\ldots(5.18)$$

A more detailed discussion (see Bopp 1959) shows that the p_{nn} must satisfy the condition

$$p_{nn} \leqslant 2[N(N-1)]^{-1}, \qquad \ldots\ldots(5.19)$$

so that the choice (5.17) will lead to a lower bound for the energy as the lowest $\frac{1}{2}N(N-1)$ pair-levels are 'over-occupied' and the higher pair-levels 'under-occupied'. We get thus the following relation for the energy

$$E \geqslant (N-1)^{-1} \sum{}' \epsilon_n, \qquad \ldots\ldots(5.20)$$

where the prime on the summation sign indicates that only the lowest $\frac{1}{2}N(N-1)$ levels are involved.

The next step in Bopp's argument is the use of experimentally known two-body energy levels to evaluate E for specially chosen atoms. For a nucleus with charge Ze surrounded by N electrons, the operator $\hat{h}$ is given by the equation

$$\hat{h} = -\frac{\hbar^2}{2m}(\nabla_1^2 + \nabla_2^2) - Ze^2\left(\frac{1}{r_1} + \frac{1}{r_2}\right) + \frac{(N-1)e^2}{r_{12}}. \qquad \ldots\ldots(5.21)$$

Changing the variables by the substitution

$$\mathbf{x}_i' = (N-1)^{-1}\mathbf{x}_i, \quad Z' = (N-1)^{-1}Z, \qquad \ldots\ldots(5.22)$$

we get

$$\hat{h} = (N-1)^2\left[-\frac{\hbar^2}{2m}(\nabla_1'^2 + \nabla_2'^2) - Z'e^2\left(\frac{1}{r_1'} + \frac{1}{r_2'}\right) + \frac{e^2}{r_{12}'}\right], \qquad \ldots\ldots(5.23)$$

which, apart from the factor $(N-1)^2$, is the helium atom problem with a nuclear charge Z'. The energy levels of $\hat{h}$ are thus known for integral values of Z' and we can use this fact to evaluate E for appropriately selected ions. For instance: Be^+ has $Z = 4$, $N = 3$, so that $Z' = 2$, and

$$E = 2\sum_{i=1}^{3}\epsilon_i(\mathrm{He}).$$

This leads to a value of 3 171 038 cm^{-1} for E which is less than 1% larger than the observed value. For the case of O^{5+} we have $Z = 8$, $N = 3$, $Z' = 4$ so that

$$E = 4\sum_{i=1}^{3}\epsilon_i(\mathrm{Be}^{2+})$$

which leads to $E = 14\,107\,380$ cm^{-1} as against the observed value of 14 104 970 cm^{-1}.

We turn now to a discussion of the Hartree–Fock and Thomas–Fermi theories of atoms (and nuclei). The density matrix is especially suitable to discuss the Thomas–Fermi theory, as it is a statistical model. In it one assumes the particles to be independent, but moving within a ' constant ' potential, the value of which is adjusted self-consistently using Poisson's equation. In the Hartree–Fock theory one tries to describe the system by ascribing to each particle a one-particle wave function. This wave function is obtained by writing down for each particle a Schrödinger equation with a potential energy which is due to all the other particles. The wave functions are varied until the whole set of equations is self-consistent. In this case we are thus also essentially describing the system in terms of independent particles, or rather, quasi-particles. If to each particle we can assign a one-particle wave function, the wave function of the whole N-particle system will be of the form

$$\Psi = C\sum_P \eta^P \prod_{i=1}^{N}\psi_{k_i}(Pi) \qquad \ldots\ldots(5.24)$$

where the ψ_{k_i} indicates a wave function corresponding to a one-particle state k_i which is a member of a complete orthonormal set, i stands for all coordinates of the ith particle, Pi for a permutation of the N i, η is $+1$ or -1 according to whether

we are dealing with bosons or fermions, the summation is over all possible permutations of the N i, and C is a normalization factor which is given by the equation

$$C^{-2} = N!, \qquad \ldots\ldots(5.25)$$

if all states k_i ($i = 1, \ldots, N$) are different. In that case we can use Eqn (5.7) to find for the one-particle density matrix

$$\langle \mathbf{x} | \hat{\rho}^{(1)} | \mathbf{x}' \rangle = N^{-1} \sum_{i=1}^{N} \psi_{k_i}{}^{*}(\mathbf{x}') \psi_{k_i}(\mathbf{x}) \qquad \ldots\ldots(5.26)$$

where we have used the orthonormality of the ψ_k. From Eqn (5.26) we find easily that

$$\begin{aligned} \langle \mathbf{x} | [\hat{\rho}^{(1)}]^2 | \mathbf{x}' \rangle &= \int \langle \mathbf{x} | \hat{\rho}^{(1)} | \mathbf{x}'' \rangle \, d^3\mathbf{x}'' \langle \mathbf{x}'' | \hat{\rho}^{(1)} | \mathbf{x}' \rangle \\ &= N^{-1} \langle \mathbf{x} | \hat{\rho}^{(1)} | \mathbf{x}' \rangle, \end{aligned}$$

or

$$[\hat{\rho}^{(1)}]^2 = N^{-1} \hat{\rho}^{(1)}. \qquad \ldots\ldots(5.27)$$

We notice that this result is *independent of the statistics*, but is a consequence solely of the fact that all k_i were different. The physical significance of Eqn (5.27) is clear: it means that $\hat{\rho}^{(1)}$ has eigenvalues 0 and N^{-1}, that is, that there are N states which each have an equal probability of being occupied, and that the other one-particle states are not occupied. In as far as the only possible wave function (5.24) for a many-fermion system is the one where all the k_i are different (Slater determinant) while there are other wave functions possible for many-boson systems, there is a certain amount of justification for calling Eqn (5.27) a Pauli condition (for instance, Young and March 1960). It is rather unfortunate that in the literature of quantum chemistry the reduced density matrices $\hat{\rho}_{\mathrm{QC}}{}^{(1)}$ have become the normally used ones. They are related to our $\hat{\rho}^{(1)}$ by the relation

$$\hat{\rho}_{\mathrm{QC}}{}^{(1)} = N\hat{\rho}^{(1)}, \qquad \ldots\ldots(5.28)$$

so that they satisfy the equations

$$\mathrm{Tr}\, \hat{\rho}_{\mathrm{QC}}{}^{(1)} = N, \quad [\hat{\rho}_{\mathrm{QC}}{}^{(1)}]^2 = \hat{\rho}_{\mathrm{QC}}{}^{(1)}, \qquad \ldots\ldots(5.29)$$

and the latter condition is sometimes wrongly interpreted as indicating a pure state. The N-body system is, of course, in a pure state, but not the one-particle systems described by the one-particle density matrix. It is correct to say that Eqn (5.27) is the necessary condition that Ψ can be described by Eqn (5.24) with all k_i different, or for the special case of fermions that the system can be described by a Slater determinant. It is interesting to note that the possible confusion between $\hat{\rho}^{(1)}$ and $\hat{\rho}_{\mathrm{QC}}{}^{(1)}$ can be traced back to Dirac's original papers (1929, 1930 a, b, 1931; for a discussion of this point see ter Haar 1960).

From Eqn (5.8) we get for the two-particle density matrix, again assuming all k_i to be different,

$$\begin{aligned} N(N-1) \langle \mathbf{x}_1, \mathbf{x}_2 | \hat{\rho}^{(2)} | \mathbf{x}_1', \mathbf{x}_2' \rangle &= \langle \mathbf{x}_1 | \hat{\rho}^{(1)} | \mathbf{x}_1' \rangle \langle \mathbf{x}_2 | \hat{\rho}^{(1)} | \mathbf{x}_2' \rangle \\ &\quad + \eta \langle \mathbf{x}_1 | \hat{\rho}^{(1)} | \mathbf{x}_2' \rangle \langle \mathbf{x}_2 | \hat{\rho}^{(1)} | \mathbf{x}_1' \rangle. \qquad \ldots\ldots(5.30) \end{aligned}$$

We note first of all that $\hat{\rho}^{(2)}$ possesses the correct symmetry properties. This follows also immediately from Eqn (5.8). Secondly, we note that for the case of

the wave function (5.24) the two-particle density matrix (and, indeed, all other many-particle density matrices) can be expressed in terms of the one-particle density matrix. The quantum-chemical two-particle density matrix $\hat{\rho}_{QC}^{(2)}$ is given by the equation

$$\hat{\rho}_{QC}^{(2)} = \tfrac{1}{2}N(N-1)\hat{\rho}^{(2)}. \qquad \ldots\ldots(5.31)$$

We note one more property of the single-particle and two-particle density matrices. Their diagonal matrices have the following physical meaning. From Eqn (5.26) it follows that

$$\langle \mathbf{x}|\hat{\rho}^{(1)}|\mathbf{x}\rangle = N^{-1}\Sigma_i|\psi_{k_i}(\mathbf{x})|^2, \qquad \ldots\ldots(5.32)$$

and we see that $\langle \mathbf{x}|\hat{\rho}^{(1)}|\mathbf{x}\rangle$ is just the normalized particle density in the system. Similarly $\langle \mathbf{x}_1, \mathbf{x}_2|\hat{\rho}^{(2)}|\mathbf{x}_1, \mathbf{x}_2\rangle$ denotes the probability density for finding simultaneously a particle at $\mathbf{x}_1$ and a particle at $\mathbf{x}_2$.

We shall from this point onwards in this section restrict our discussion to fermion systems ($\eta = -1$) where the wave function (5.24) corresponds to a Slater determinant. Moreover, we shall be interested in one particular state of the N-particle system, namely the ground state. We have seen earlier in this section that the energy can be expressed in terms of the two-particle density matrix and one method of approach is by considering possible expressions for $\hat{\rho}^{(2)}$. This method was first suggested by Mayer (1955) and has since then been explored by various authors (Koppe 1957, Tredgold 1957, Mizuno and Isuyama 1957, Young and March 1960). On the other hand, we have also seen that for a Slater determinant the two-particle density matrix can be expressed in terms of the one-particle density matrix and we might try to find a suitable expression for $\hat{\rho}^{(1)}$ (see Golden 1957 a, b, Naqvi 1959). We shall briefly consider both methods.

From Eqns (5.10) and (2.9) we see that we have for the energy of the system

$$E = \tfrac{1}{2}N \operatorname{Tr} \hat{h}\hat{\rho}^{(2)}, \qquad \ldots\ldots(5.33)$$

where $\hat{h}$ is given by Eqn (5.11) while $\hat{\rho}^{(2)}$ must satisfy the normalization condition

$$\operatorname{Tr} \hat{\rho}^{(2)} = 1 \qquad \ldots\ldots(5.34)$$

and the antisymmetry condition

$$\langle \mathbf{x}_1, \mathbf{x}_2|\hat{\rho}^{(2)}|\mathbf{x}_1', \mathbf{x}_2'\rangle = -\langle \mathbf{x}_1, \mathbf{x}_2|\hat{\rho}^{(2)}|\mathbf{x}_2', \mathbf{x}_1'\rangle = -\langle \mathbf{x}_2, \mathbf{x}_1|\hat{\rho}^{(2)}|\mathbf{x}_1', \mathbf{x}_2'\rangle. \qquad \ldots\ldots(5.35)$$

Taking care to see that Eqns (5.35) are satisfied at all times one can use Eqns (5.33) and (5.34) for a variational method to evaluate $\hat{\rho}^{(2)}$ and then E.

In the other method we consider only $\hat{\rho}^{(1)}$. From Eqns (5.9) and (5.30) we have in terms of $\hat{\rho}^{(1)}$

$$E = \operatorname{Tr} \hat{H}\hat{\rho}^{(1)}, \qquad \ldots\ldots(5.36)$$

where

$$\langle \mathbf{x}|\hat{H}|\mathbf{x}'\rangle = N\langle \mathbf{x}|\hat{H}_1|\mathbf{x}'\rangle + \tfrac{1}{2}\langle \mathbf{x}|\hat{U}|\mathbf{x}'\rangle, \qquad \ldots\ldots(5.37)$$

$$\langle \mathbf{x}|\hat{U}|\mathbf{x}'\rangle = \int \langle \mathbf{x}|\hat{V}_{12}|\mathbf{x}''\rangle d^3\mathbf{x}''\langle \mathbf{x}''|\hat{\rho}^{(1)}|\mathbf{x}'\rangle \delta(\mathbf{x}-\mathbf{x}') - \langle \mathbf{x}|\hat{V}_{12}|\mathbf{x}'\rangle\langle \mathbf{x}|\hat{\rho}^{(1)}|\mathbf{x}'\rangle. \qquad \ldots\ldots(5.38)$$

If we vary expression (5.36) with respect to $\hat{\rho}^{(1)}$ and bear in mind the condition (5.27) on $\hat{\rho}^{(1)}$ and the fact that $\hat{H}$ contains $\hat{\rho}^{(1)}$ through $\hat{U}$, we find the following equation to be satisfied by $\hat{\rho}^{(1)}$,

$$[\hat{H}', \hat{\rho}^{(1)}]_- = 0, \qquad \ldots\ldots(5.39)$$

with

$$\hat{H}' = N\hat{H}_1 + \hat{U}. \qquad \ldots\ldots(5.40)$$

If we had varied the ψ_k which occur in the Slater determinant instead of $\hat{\rho}^{(1)}$, we would have found for the ψ_k the Schrödinger equation

$$\hat{H}'\psi_k = \epsilon_k \psi_k, \qquad \ldots\ldots(5.41)$$

which is just the Hartree–Fock equation.

In the case of the Thomas–Fermi theory we are interested in very large systems. If the system were infinite, $\langle \mathbf{x}|\hat{\rho}^{(1)}|\mathbf{x}'\rangle$ should depend on $\mathbf{x}-\mathbf{x}'$ only, and we may assume that this is still correct for the actual systems studied. Moreover, one assumes in the Thomas–Fermi method that N is very large. If N were infinite, the off-diagonal elements of $\hat{\rho}^{(1)}$ would be zero (this follows from Eqn (5.26), and the fact that we would then sum over the complete set: see, for instance, Kramers 1957, p. 129). It is thus plausible to assume that for the actual system, the off-diagonal elements of $\hat{\rho}^{(1)}$ will be small and can be expanded in a Taylor series in $\mathbf{x}-\mathbf{x}'$. The method to be used now is the following one. We replace the exact Hamiltonian by a model Hamiltonian $\hat{H}_M$ (compare the use of a harmonic oscillator well in shell model discussions), and require $\hat{\rho}^{(1)}$ to satisfy the following conditions:

(i) Eqn (5.27), (ii) the normalization condition (2.9),

(iii)
$$[\hat{H}_M, \hat{\rho}^{(1)}]_- = 0, \qquad \ldots\ldots(5.42)$$

and (iv) the requirement that the model energy E_M given by the equation

$$E_M = \mathrm{Tr}\,\hat{H}_M\hat{\rho}^{(1)} \qquad \ldots\ldots(5.43)$$

be a minimum. Once we have found $\hat{\rho}^{(1)}$ the total energy is given by Eqn (5.36).

To a first approximation we can use for the ψ_k in Eqn (5.26) plane waves. We then get from conditions (i) to (iv) an energy density $\epsilon(\mathbf{x})$ which is given by

$$\epsilon = \frac{\hbar^2}{8m}\frac{12}{5}(6\pi^2)^{2/3}n^{5/3} + nV_M(\mathbf{x}), \qquad \ldots\ldots(5.44)$$

where V_M is the potential energy in the model Hamiltonian (assumed to a first approximation to be a constant) and n the particle density,

$$n(\mathbf{x}) = N_0\langle \mathbf{x}|\hat{\rho}^{(1)}|\mathbf{x}\rangle, \qquad \ldots\ldots(5.45)$$

where N_0 is the number of particles per unit volume. In the next approximation one uses the fact that N is finite and one introduces the variation with $\mathbf{x}$ of both V_M and n, and requires the total energy $\int\epsilon(\mathbf{x})\,d^3\mathbf{x}$ to be a minimum. The result is then the following equation for n

$$\frac{\hbar^2}{8m}\left[\frac{(\nabla n)^2}{n^2} - 2\frac{\nabla^2 n}{n}\right] + \frac{\hbar^2}{2m}(6\pi^2)^{2/3}n^{5/3} + V(\mathbf{x}) = \epsilon_0, \qquad \ldots\ldots(5.46)$$

where ϵ_0 is the energy of the most energetic particle in the system. Eqn (5.46) is the so-called von Weizsäcker equation (1935). We refer to Naqvi's paper (1959) for details.

§ 6. The Density Matrix in Solid State Physics

Both density matrix and Green function techniques have been used in recent years to discuss many solid state problems, both equilibrium and non-equilibrium problems. In the present section we shall be concerned with equilibrium properties and in the next section with non-equilibrium problems. Once again we can only discuss a few selected topics and must refer to the literature for a more comprehensive coverage. For this section we may mention the work on many-electron systems (Zubarev 1954, Mayer 1955, Koppe 1957, Mizuno and Izuyama 1957, Tredgold 1957, Edwards 1958 a, Ehrenreich and Cohen 1959, Goldstone and Gottfried 1959, Kanazawa and Watabe 1960, Kanazawa, Misawa and Fujita 1960, Young and March 1960), on diamagnetism of a system of free electrons (Sondheimer and Wilson 1951, Nakajima 1955, Nitsovich 1959, Kanazawa and Matsudaira 1960), on diamagnetism of conduction electrons (Nakajima 1956, Enz 1960 a, b, Hebborn and Sondheimer 1960), on electron–phonon systems (Matsubara 1955 a, b, Nakajima 1955, Ichimura 1956), on semiconductors (Bonch-Bruevich 1956 c, Krivoglaz and Pekar 1957 a, b, Osaka 1959), on superconductors (Salam 1953, Schafroth, Butler and Blatt 1957, Koppe and Mühlschlegel 1958, Valatin 1958), on ferromagnetism (Bogolyubov and Tyablikov 1959, Brout 1959), and on surface tension in liquids (Ono 1958). In this section we shall restrict ourselves to a discussion of the diamagnetic susceptibility of a system of electrons, including a brief discussion of Schafroth's treatment (1951) of the Meissner effect.

In discussing the diamagnetic susceptibility we are dealing with an equilibrium aspect and one could do this by evaluating the partition function. We saw in § 4 (Eqns (4.4) and (4.5)) that this can be done by finding the trace of $e^{-\beta\hat{H}}$. This is usually done by finding all energy levels of the system and writing (we restrict ourselves here to *petit ensembles*)

$$Z = \Sigma_n \exp(-\beta E_n), \qquad \ldots\ldots(6.1)$$

where the summation is over all energy levels of the system. There is, however, no reason whatever to restrict oneself to this particular form of the trace. If we introduce once again the operator $\hat{\rho}(\beta)$ by the equation

$$\hat{\rho}(\beta) = e^{-\beta\hat{H}}, \qquad \ldots\ldots(6.2)$$

which satisfies the Bloch equation (4.6),

$$\partial\hat{\rho}/\partial\beta = -\hat{H}\hat{\rho}, \qquad \ldots\ldots(6.3)$$

we have

$$Z = \mathrm{Tr}\,\hat{\rho}(\beta). \qquad \ldots\ldots(6.4)$$

The magnetic susceptibility χ follows from the usual formula

$$\chi = (\beta\mathscr{H}Z)^{-1}\partial Z/\partial\mathscr{H}, \qquad \ldots\ldots(6.5)$$

where $\mathscr{H}$ is the magnetic field, which we shall assume to be uniform.

We shall assume the particles (electrons) in our system to be independent, so that the many-particle partition function is the Nth power of the single-particle partition function. If we take χ to be the susceptibility per particle, we can still use Eqns (6.2) to (6.5), but with $\hat{H}$ being the single-particle Hamiltonian which is of the form

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 - \frac{ie\hbar}{mc}(\mathbf{A}.\nabla) + \frac{e^2\mathbf{A}^2}{2mc^2}, \qquad \ldots\ldots(6.6)$$

where $\mathbf{A}$ is the vector potential,

$$\mathbf{A} = \tfrac{1}{2}[\mathscr{H} \wedge \mathbf{x}], \qquad \ldots\ldots(6.7)$$

$-e$ the charge of the electron, and m the electron mass. It will be noted that we are not introducing the electron spin and are thus restricting the discussion to Boltzmann statistics. We refer to the paper by Sondheimer and Wilson (1951) for a discussion of the Fermi–Dirac case (we may also refer to the discussion in § 9.4 of ter Haar 1954; the $W_{N,\mathrm{Bo}}$ of that section is related to the diagonal element of the $\langle \mathbf{x}|\hat{p}(\beta)|\mathbf{x}'\rangle$ we have here).

To evaluate $\hat{p}(\beta)$, it is convenient to change to the coordinate representation. As was noted in § 2, any complete orthonormal set $\varphi_n(\mathbf{x})$ can be used, and it can be shown that (compare Eqn (2.48))

$$\langle \mathbf{x}|\hat{p}(\beta)|\mathbf{x}'\rangle = \textstyle\sum_m \varphi_m^*(\mathbf{x}') \exp(-\beta \hat{H}) \varphi_m(\mathbf{x}), \qquad \ldots\ldots(6.8)$$

where $\hat{H}$ operates on $\varphi_m(\mathbf{x})$ only. From Eqn (6.8) and the orthonormality of the set φ_n it follows that

$$\langle \mathbf{x}|\hat{p}(0)|\mathbf{x}'\rangle = \textstyle\sum_m \varphi_m^*(\mathbf{x}') \varphi_m(\mathbf{x}) = \delta(\mathbf{x}-\mathbf{x}'). \qquad \ldots\ldots(6.9)$$

We must thus find a solution of Eqn (6.3) which satisfies the boundary condition (6.9). We shall assume the magnetic field to be along the z axis so that we can write Eqn (6.3), using Eqns (6.6) and (6.7), in the form

$$\frac{\partial \hat{p}}{\partial \beta} = \left[\frac{\hbar^2}{2m}\nabla^2 - \frac{e\hbar\mathscr{H}}{2imc}\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) - \frac{e^2\mathscr{H}^2}{8m^2c^2}(x^2+y^2)\right]\hat{p}. \qquad \ldots\ldots(6.10)$$

If $\mathscr{H} = 0$, Eqn (6.10) reduces to the heat conduction equation with the solution (for instance, Carslaw and Jaeger 1947)

$$\hat{p}_{\mathscr{H}=0} = \nu^{3/2} \exp[-\pi\nu(\mathbf{x}-\mathbf{x}'.\mathbf{x}-\mathbf{x}')], \quad \nu = m/2\pi\hbar^2\beta. \qquad \ldots\ldots(6.11)$$

If $\mathscr{H} \neq 0$, one finds a more complicated expression (see Sondheimer and Wilson (1951) for a derivation of this equation)

$$\langle \mathbf{x}|\hat{p}(\beta)|\mathbf{x}'\rangle = f(\beta)\exp[-(ie\mathscr{H}/2\hbar c)(x'y-y'x) - g(\beta)\{(x-x')^2+(y-y')^2\} - \pi\nu(z-z')^2], \qquad \ldots\ldots(6.12)$$

where $f(\beta)$ and $g(\beta)$ are given by the expressions

$$g = \frac{e\mathscr{H}}{4\hbar c}\coth\frac{e\hbar\mathscr{H}\beta}{2mc}, \quad f = \nu^{1/2}\frac{e\mathscr{H}}{2\hbar c}\operatorname{cosech}\frac{e\hbar\mathscr{H}\beta}{2mc}. \qquad \ldots\ldots(6.13)$$

The partition function per unit volume is given by

$$Z = V^{-1}\int \langle \mathbf{x}|\hat{p}(\beta)|\mathbf{x}\rangle\, d^3\mathbf{x}, \qquad \ldots\ldots(6.14)$$

and from Eqns (6.12), (6.13) and (6.14) we get

$$Z = \nu^{3/2}[\gamma/\sinh\gamma], \quad \gamma = \beta\mu_B\mathscr{H}, \qquad \ldots\ldots(6.15)$$

where $\mu_B = e\hbar/2mc$ is the Bohr magneton. Eqn (6.15) leads to the following expression for χ, if we use Eqn (6.5),

$$\chi = -\mu_B(\coth\gamma - \gamma^{-1}), \qquad \ldots\ldots(6.16)$$

which for weak fields reduces to the Landau expression (1930)

$$\chi = -\tfrac{1}{3}\beta\mu_B{}^2. \qquad \ldots\ldots(6.17)$$

If we are dealing with conduction electrons, we must replace the Hamiltonian of Eqn (6.6) by

$$\hat{H}' = \hat{H} + \hat{V}, \qquad \ldots\ldots(6.18)$$

where $\hat{V}$ is the periodic lattice potential. It is no longer possible to evaluate $\langle \mathbf{x}|\hat{p}(\beta)|\mathbf{x}'\rangle$ exactly, but one uses an expression in powers of $\mathscr{H}$, the term in $\mathscr{H}^2$ giving the field-independent part of the susceptibility (Hebborn and Sondheimer 1960).

To conclude this section we consider briefly the related question of the Meissner effect, that is, the exclusion of a steady magnetic field from a superconductor. We consider the external magnetic field as a perturbation, and write for the Hamiltonian

$$\hat{H} = \hat{H}_0 + \hat{H}_1, \qquad \ldots\ldots(6.19)$$

where $\hat{H}_0$ is the kinetic energy and $\hat{H}_1$ given by the equation (we are now considering a system which is not necessarily uniform and $\hat{H}$ is once again the Hamiltonian of the total system)

$$\hat{H}_1 = -c^{-1}\int (\mathbf{A}(\mathbf{x}).\mathbf{j}(\mathbf{x}))\, d^3\mathbf{x}, \qquad \ldots\ldots(6.20)$$

where $\mathbf{j}$ is the total current density given by the expression :

$$\mathbf{j} = \frac{e\hbar}{2mi}(\psi^* \nabla\psi - \psi\nabla\psi^*) - \frac{e^2}{mc}\psi^*\psi\mathbf{A}. \qquad \ldots\ldots(6.21)$$

The components $J_\mu(\mathbf{x})$ of the average current density are given by the equations

$$J_\mu(\mathbf{x}) = \langle j_\mu(\mathbf{x})\rangle. \qquad \ldots\ldots(6.22)$$

As this average is evaluated using the equilibrium density matrix (4.2) we get to a first approximation for J_μ a linear expression in the components A_ν of the vector potential

$$J_\mu(\mathbf{x}) = \Sigma_\nu \int K_{\mu\nu}(\mathbf{x}-\mathbf{x}')\, A_\nu(\mathbf{x}')\, d^3\mathbf{x}'. \qquad \ldots\ldots(6.23)$$

From Eqns (6.20) and (6.22) and the expansion

$$\exp(-\beta\hat{H}) \simeq \exp(-\beta\hat{H}_0)\,[1-\beta\hat{H}_1] \qquad \ldots\ldots(6.24)$$

it follows that the $K_{\mu\nu}$ are proportional to the correlation functions $\phi_{\mu\nu}(\mathbf{x}-\mathbf{x}')$ given by the equation

$$\phi_{\mu\nu}(\mathbf{x}-\mathbf{x}') = \langle j_\nu(\mathbf{x}')\, j_\mu(\mathbf{x})\rangle_0, \qquad \ldots\ldots(6.25)$$

where the average is taken over the unperturbed equilibrium ensemble, that is, the ensemble with $\hat{p}(\beta) = \exp(-\beta\hat{H}_0)$. We see that the $\phi_{\mu\nu}$ are essentially an example of the correlation functions (3.14).

The behaviour of the Fourier transforms of the $K_{\mu\nu}$ determines the diamagnetic properties of the system. From general symmetry considerations it follows that the $\mu\nu$-dependence of the Fourier transforms $K_{\mu\nu}(\mathbf{k})$ can be expressed as follows

$$K_{\mu\nu}(\mathbf{k}) = (k^2\,\delta_{\mu\nu} - k_\mu k_\nu)\, K(k^2). \qquad \ldots\ldots(6.26)$$

If $K(k^2)$ has no singularity at $k = 0$ one finds ordinary diamagnetism. If, however, $K(k^2)$ has a pole at $k = 0$,

$$K(k^2) = \alpha k^{-2} + K_r(k^2), \qquad \ldots\ldots(6.27)$$

where $K_r(k^2)$ behaves regularly at $k = 0$, the term $K_r(k^2)$ leads to ordinary diamagnetism, but the first term leads to an extra term, $\mathbf{J}_s$, in $\mathbf{J}$ which can be shown to satisfy the London equation

$$[\nabla \wedge \mathbf{J}_s] = \alpha \mathbf{B}, \qquad \ldots\ldots(6.28)$$

from which the Meissner effect follows in the usual way. We see from this that the Meissner effect is related to a singularity in the Fourier transform of the kernel $K_{\mu\nu}(\mathbf{x}-\mathbf{x}')$, that is, it is related to long range correlations in the system.

§ 7. Non-equilibrium Processes; Transport Theory

In the preceding sections we have mainly been concerned with systems in equilibrium. The question arises, however, how equilibrium is reached, and what happens if we are dealing with a system which is permanently in non-equilibrium, but which has reached a steady state? Such situations are described by transport equations. The basic, exact equation of motion is Eqn (2.37), and this is the equation from which the equations discussed in the present section are derived. We shall briefly discuss the derivation of a transport equation starting from Eqn (2.37), but the main discussion in the present section will be of the evaluation of transport coefficients without the intermediary of a transport equation, a subject which has been developed very extensively in recent years and which was started by the fundamental papers of Kubo (1956, 1957 a) and Nakano (1956, 1957, 1959, 1960 a, b, c). We may refer also to papers by Callen and Welton (1951), Nakajima (1956), Kohn and Luttinger (1957), Argyres (1958, 1960 a, b), Edwards (1958 b), Greenwood (1958), Lax (1958), Luttinger (1958), Luttinger and Kohn (1958), Mattis and Bardeen (1958), Adams and Holstein (1959), Argyres and Roth (1959), Bonch-Bruevich (1959 e), Chester and Thellung (1959), Klinger (1959 a, b, c, d, e), Konstantinov and Perel' (1959, 1960), Kubo, Hasegawa and Hashitsuma (1959), McLennan (1959), Montroll and Ward (1959), Zigenlaub (1959), Green (1960), Kanazawa and Watabe (1960), Yokota (1960) and Zubarev (1960). Among the papers dealing with the more general problem of irreversible processes and transport equations we refer to those by Born and Green (1946, 1947 a, b, 1948), Green (1947), Mori and Ono (1952), Ono (1953, 1954, 1958), Ross and Kirkwood (1954), Van Hove (1955), Kümmel (1955), van Kampen (1956), Mori (1956, 1958, 1959), Gurzhi (1957), Husimi, Kitano and Nishiyama (1958), Klimontovich and Temko (1958), Prigogine and Balescu (1959 a, b), Prigogine and Ono (1959), Helfand (1960), Helfand and Rice (1960), Klimontovich and Silin (1960), Nishikawa (1960), von Roos (1960), Saitô (1960), and Żelazny (1960). We also refer to the Proceedings of the 1959 Varenna Summer School (Varenna 1959) and the 1956 Brussels Conference (Prigogine 1958). Finally, we refer to the discussion in § 9 where we shall consider relaxation processes.

In classical statistical mechanics the counterpart of the equation of motion for the density matrix

$$i\hbar\dot{\hat{\rho}} = [\hat{H}, \hat{\rho}]_- \qquad \ldots\ldots(7.1)$$

is the Liouville equation. Both in Eqn (7.1) and in the Liouville equation we are dealing with the total N-body system. The ordinary transport equation, on the other hand, involves, if possible, only quantities referring to a single particle. In classical theory the transport equation is obtained by integrating over the coordinates of all but one of the particles. As the Hamiltonian occurring in the Liouville equation will in general contain terms referring to two particles (two-body interaction terms), the transport equation for the single-particle distribution function will contain the two-body distribution function. One can either make some simplifying assumptions about this two-body distribution function, expressing it in terms of the single-particle distribution function, or one can set up a transport equation for the two-body distribution function by integrating the Liouville equation over the coordinates of $N-2$ particles only. The resultant equation will, however, contain the three-body distribution function, and we have to continue the process. We may refer here to a similar situation which we met in § 3 with respect to the Green functions.

In quantum mechanics a similar situation arises, but there is an additional complication in that the density matrix $\hat{\rho}$ occurring in Eqn (7.1) is not the immediate counterpart of the N-body distribution function $f(\mathbf{x}_1, ..., \mathbf{x}_N\,;\ \mathbf{p}_1, ..., \mathbf{p}_N)$. The latter is a function of N coordinates $\mathbf{x}_i$ and N momenta $\mathbf{p}_i$, while the density matrix in coordinate representation $\langle \mathbf{x}_1, ..., \mathbf{x}_N | \hat{\rho} | \mathbf{x}_1', ..., \mathbf{x}_N' \rangle$ is a function of N coordinates $\mathbf{x}_i$ and N coordinates $\mathbf{x}_i'$. In working with a transport equation, and especially in evaluating transport coefficients from a solution of the transport equation, it is often more convenient to deal with a distribution function such as the classical one. One can actually construct from the density matrix such a distribution function, the so-called Wigner distribution functions (Wigner 1932)

$$f^{\mathrm{W}}(\mathbf{x}_1, ..., \mathbf{x}_N;\ \mathbf{p}_1, ..., \mathbf{p}_N) = (2\pi)^{-3N} \int \langle \mathbf{x}_i - \tfrac{1}{2}\hbar\boldsymbol{\tau}_i | \hat{\rho} | \mathbf{x}_i + \tfrac{1}{2}\hbar\boldsymbol{\tau}_i \rangle \exp\Big\{ -i \sum_{j=1}^{N} (\boldsymbol{\tau}_j . \mathbf{p}_j) \Big\} d^{3N}\boldsymbol{\tau}. \qquad \ldots\ldots(7.2)$$

From Eqn (2.7) for the average value of a quantity $\hat{A}$,

$$\langle \hat{A} \rangle = \mathrm{Tr}\, \hat{\rho}\hat{A}, \qquad \ldots\ldots(7.3)$$

Eqn (7.2), and the well-known properties of Fourier integrals, one can prove that for any function $A(\mathbf{x}_i, \mathbf{p}_i)$ of the coordinates and momenta

$$\langle A \rangle = \int A(\mathbf{x}_i, \mathbf{p}_i) f^{\mathrm{W}}(\mathbf{x}_i;\ \mathbf{p}_i)\, d^{3N}\mathbf{x}\, d^{3N}\mathbf{p}, \qquad \ldots\ldots(7.4)$$

so that the f^{W} have, indeed, the properties of distribution functions. Moreover, in the limit as $\hbar \to 0$ (classical limit) the f^{W} go over into the classical distribution functions.

From Eqns (7.1) and (7.2) it follows that f^{W} satisfies the equation of motion

$$-i\hbar \dot{f}^{\mathrm{W}}(\mathbf{x}_i, \mathbf{p}_i) = (2\pi)^{-6N} \int \{ H(\mathbf{q}_i + \tfrac{1}{2}\hbar\mathbf{k}_i, \mathbf{y}_i - \tfrac{1}{2}\hbar\boldsymbol{\tau}_i) - H(\mathbf{q}_i - \tfrac{1}{2}\hbar\mathbf{k}_i, \mathbf{y}_i + \tfrac{1}{2}\hbar\boldsymbol{\tau}_i) \}$$
$$\times f^{\mathrm{W}}(\mathbf{y}_i, \mathbf{q}_i) \exp i \textstyle\sum_j [(\boldsymbol{\tau}_j . \mathbf{q}_j - \mathbf{p}_j) + (\mathbf{k}_j . \mathbf{y}_j - \mathbf{x}_j)]\, d^{3N}\boldsymbol{\tau}\, d^{3N}\mathbf{k}\, d^{3N}\mathbf{q}\, d^{3N}\mathbf{y}, \qquad \ldots\ldots(7.5)$$

where $H(\mathbf{x}_i, \mathbf{p}_i)$ is the Hamiltonian of the system. As $\hbar \to 0$, the equation tends to the Liouville equation.

By integrating f^{W} over the arguments of all particles, but one (or two, ...) we obtain a one-particle (two-body, ...) distribution function, and by similar integrations of Eqn (7.5) we obtain the transport equations for the quantum mechanical (Wigner) distribution functions. The situation is now completely similar to the one in classical statistics.

In view of the fact that the transport equation for the single-particle distribution function contains either the two-body distribution function or involves simplifying assumptions, many authors have discussed the possibility of obtaining transport coefficients without solving a transport equation. In all these papers one assumes—as is usually done in a discussion of the transport equation—that the perturbing field, density gradient, temperature gradient, ... are so small that one can restrict the discussion to an approximation which is linear in the perturbation. If this is the case, one can show that the transport coefficients, that is, typical non-equilibrium quantities, can be evaluated using an equilibrium density matrix. In fact, such properties as electrical conductivity are directly related to the response of a system to an external force, and Callen and Welton (1951) have shown that this response in turn is related through Nyquist's fluctuation-dissipation theorem to time correlation functions, that is, to Green functions (see § 3). Of course, in as far as one uses the equation of motion (7.1) for the density matrix to obtain the expressions for the transport coefficients, one has used a transport equation, but only one for an N-body system, not the one for the one-particle distribution function. One can show that the expressions obtained for the transport coefficients are the same as those obtained from the usual transport equation in the appropriate limits of sufficiently weak perturbations.

To derive an expression for the electrical conductivity we first of all consider the general problem of the response of a system to a time-dependent, oscillating perturbation which is switched on adiabatically. That is, we assume the total Hamiltonian to be of the form

$$\hat{H} = \hat{H}_0 + \hat{H}'(t), \qquad \ldots\ldots(7.6)$$

where

$$\hat{H}'(-\infty) = 0, \quad \hat{H}'(t) = e^{\epsilon t} e^{i\omega t} \hat{V} \quad (\epsilon \to 0, \epsilon > 0), \qquad \ldots\ldots(7.7)$$

with $\hat{V}$ a time-independent operator and ω the frequency of the perturbation. We assume the density matrix at $t = -\infty$ to be the equilibrium density matrix

$$\hat{\rho}(-\infty) = \hat{\rho}_0 = Z^{-1} \exp(-\beta \hat{H}_0), \qquad \ldots\ldots(7.8)$$

where Z is the partition function : $Z = \mathrm{Tr}[\exp(-\beta \hat{H}_0)]$. For the density matrix at a later time we write

$$\hat{\rho}(t) = \hat{\rho}_0 + \hat{\Delta\rho}, \qquad \ldots\ldots(7.9)$$

and from Eqns (7.1), (7.6) and (7.9) we get, if we neglect second-order terms,

$$i\hbar \hat{\Delta\rho} = [\hat{H}_0, \hat{\Delta\rho}]_- + [\hat{H}', \rho_0]_-. \qquad \ldots\ldots(7.10)$$

Introducing the Heisenberg operators

$$\hat{\Omega}^{H} = \exp(i\hat{H}_0 t/\hbar)\, \hat{\Omega} \exp(-i\hat{H}_0 t/\hbar), \qquad \ldots\ldots(7.11)$$

we have

$$i\hbar \hat{\Delta\rho}^{H} = \exp(i\hat{H}_0 t/\hbar)[\hat{H}', \hat{\rho}_0]_- \exp(-i\hat{H}_0 t/\hbar), \qquad \ldots\ldots(7.12)$$

with the solution

$$\widehat{\Delta\rho} = (i\hbar)^{-1}\int_{-\infty}^{t} \exp\left[i\hat{H}_0(\tau-t)/\hbar\right]\left[\hat{H}'(\tau), \hat{\rho}_0\right]_- \exp\left[-i\hat{H}_0(\tau-t)/\hbar\right] d\tau. \quad \ldots\ldots(7.13)$$

From Eqns (7.3), (7.9) and (7.13) we get for the average value $\langle \hat{A} \rangle$,

$$\langle \hat{A}(t) \rangle = \langle \hat{A} \rangle_0 + (i\hbar)^{-1}\int_{-\infty}^{t} \langle [\hat{A}(t), \hat{V}(\tau)]_- \rangle_0 \, e^{i\omega\tau + \varepsilon\tau} \, d\tau, \quad \ldots\ldots(7.14)$$

where $\langle \ldots \rangle_0$ indicates an average over the equilibrium ensemble $\hat{\rho}_0$.

If we apply this to the case of the electrical conductivity, $\hat{A}$ is replaced by one of the components $\hat{\jmath}_\mu$ of the current and $\hat{H}'$ is of the form

$$\hat{H}' = -e\sum_j (\mathbf{E}.\mathbf{x}_j)\, e^{i\omega t} \quad \ldots\ldots(7.15)$$

where we have now dropped the adiabatic factor $e^{\varepsilon t}$ and where $\mathbf{E}$ is the amplitude of the electrical field strength. We shall apply Eqn (7.14) to the case $\hat{A} = \hat{\jmath}_\mu$, rewriting it slightly, using the fact that $\langle \hat{\jmath}_\mu \rangle_0 = 0$, and using Eqn (7.11) for $\hat{\jmath}_\mu{}^{\mathrm{H}}$. The result is

$$\begin{aligned}\langle j_\mu \rangle &= (i/\hbar)\,\mathrm{Tr}\int_{-\infty}^{t} \exp\left[i\hat{H}_0(\tau-t)/\hbar\right]\left[e\sum_j (\mathbf{E}.\mathbf{x}_j), \hat{\rho}_0\right]_- \exp\left[-i\hat{H}_0(\tau-t)/\hbar\right] e^{i\omega\tau}\hat{\jmath}_\mu \, d\tau \\ &= (i/\hbar)\,\mathrm{Tr}\int_{-\infty}^{t} \left[e\sum_j (\mathbf{E}.\mathbf{x}_j), \hat{\rho}_0\right]_- \hat{\jmath}_\mu{}^{\mathrm{H}}(t-\tau)\, e^{i\omega\tau}\, d\tau \\ &= (i/\hbar)\,\mathrm{Tr}\int_{0}^{\infty} \left[e\sum_j (\mathbf{E}.\mathbf{x}_j), \hat{\rho}_0\right]_- \hat{\jmath}_\mu{}^{\mathrm{H}}(\tau)\, e^{i\omega(t-\tau)}\, d\tau. \quad \ldots\ldots(7.16)\end{aligned}$$

The complex conductivity tensor, which is defined by the relation

$$\langle \hat{\jmath}_\mu \rangle = \sigma_{\mu\nu} E_\nu e^{i\omega t}, \quad \ldots\ldots(7.17)$$

follows from Eqn (7.16) and is equal to

$$\sigma_{\mu\nu} = (i/\hbar)\,\mathrm{Tr}\int_{0}^{\infty} e^{-i\omega\tau}\left[e\sum_j x_{j\nu}, \hat{\rho}_0\right]_- \hat{\jmath}_\mu{}^{\mathrm{H}}(\tau)\, d\tau. \quad \ldots\ldots(7.18)$$

We shall use the identity

$$\left[\hat{\Omega}, \exp(-\beta\hat{H}_0)\right]_- = -i\hbar \exp(-\beta\hat{H}_0)\int_0^\beta \exp(\lambda\hat{H}_0)\,\dot{\hat{\Omega}}\exp(-\lambda\hat{H}_0)\, d\lambda, \quad \ldots\ldots(7.19)$$

where Eqn (2.38) has been used. This equation follows easily, as was observed by Kubo (1957 a) if one uses a representation in which $\hat{H}_0$ is diagonal. As $\hat{\rho}_0$ is given by Eqn (7.8), we find from Eqn (7.19)

$$\left[e\sum_j x_{j\nu}, \hat{\rho}_0\right]_- = (i\hbar/Z)\exp(-\beta\hat{H}_0)\int_0^\beta \exp(\lambda\hat{H}_0)\,\hat{\jmath}_\nu \exp(-\lambda\hat{H}_0)\, d\lambda, \quad \ldots\ldots(7.20)$$

or, using Eqns (7.18), (7.20) and (7.11),

$$\sigma_{\mu\nu} = Z^{-1}\int_0^\infty d\tau\, e^{-i\omega\tau}\int_0^\beta \langle \hat{\jmath}_\nu{}^{\mathrm{H}}(-i\hbar\lambda)\hat{\jmath}_\mu{}^{\mathrm{H}}(\tau)\rangle_0\, d\lambda, \quad \ldots\ldots(7.21)$$

and we see once again the importance of the correlation between currents (compare Eqn (6.27) of the previous section).

Equation (7.21) has been used by Montroll and Ward (1959) and Konstantinov and Perel' (1960) as the basis of diagram techniques to evaluate transport coefficients.

§ 8. Polarization, Scattering and Angular Correlation Experiments

We mentioned in the introduction that often one can use an operational approach to the density matrix. This is especially the case when one discusses polarization or scattering experiments, as one is in that case interested in only a few of the many parameters which specify the system and one can use a density matrix which refers to only those degrees of freedom which are studied experimentally. The simplest example is the polarization of a beam of electrons which we shall discuss in some detail. A related problem is that of the polarization of light which we shall also consider here, without going into a very detailed discussion. Our final discussion will be of scattering and angular correlation experiments. For a consideration of such experiments elaborate matrix techniques have been developed, some of which are based upon the density matrix, but we must refer to the literature for a discussion of such techniques (see, for instance, Tolhoek and de Groot 1951 b, c, Cox and Tolhoek 1953, Tolhoek and Cox 1953, Hartogh, Tolhoek and de Groot 1954, de Groot and Tolhoek 1955, Huby 1958). We can only touch upon some aspects of density matrix techniques; for more details of those techniques as applied to polarization, scattering, and angular correlation experiments in nuclear physics we may refer, for instance, to the papers by Tolhoek and de Groot (1951 a), Lipps and Tolhoek (1954 a, b), Kotani (1955), Tolhoek (1956), Fano (1957), Hagedorn (1958) and Zaidi (1959). We may also mention some unpublished lecture notes based upon lectures delivered by L. Rosenfeld to Nordita in Copenhagen which have been of great use in writing this section. The use of density matrices in discussing the polarization of electromagnetic radiation is based upon the fundamental ideas of Stokes (1852). Recently Fano (1949, 1957) and Wolf and Roman (see, for instance, Wolf 1954, 1959 a, b, 1960, Roman 1959, Parrent and Roman 1960) have discussed this problem in great detail, but we can refer here only to those papers which are relevant to our discussion of density matrix techniques. In the present section we shall consider the polarization of particles with spin $\frac{1}{2}$; the case of larger spin values will be mentioned only briefly.

Let us consider a beam of particles with spin $\frac{1}{2}$, for instance, a beam of electrons. If we are only interested in the polarization or spin-orientation properties of this beam, we have a system of particles with two degrees of freedom, as long as we neglect negative energy states as we shall do here. The system should thus be describable by a 2 by 2 density matrix, and we need only 3 independent parameters to determine fully the density matrix (compare the discussion in § 2). The physical situation is completely defined, if we know the polarization vector $\mathbf{P}$, that is, the average value of the spin vector in the system,

$$\mathbf{P} = \langle \boldsymbol{\sigma} \rangle, \qquad \ldots\ldots (8.1)$$

where $\hat{\boldsymbol{\sigma}}$ is the vector the components of which are the Pauli matrices,

$$\hat{\sigma}_x = \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix}, \quad \hat{\sigma}_y = \begin{Vmatrix} 0 & -i \\ i & 0 \end{Vmatrix}, \quad \hat{\sigma}_z = \begin{Vmatrix} 1 & 0 \\ 0 & -1 \end{Vmatrix}. \qquad \ldots\ldots(8.2)$$

We note that these matrices satisfy the following relations

$$\mathrm{Tr}\,\hat{\sigma}_x = \mathrm{Tr}\,\hat{\sigma}_y = \mathrm{Tr}\,\hat{\sigma}_z = 0; \qquad \ldots\ldots(8.3a)$$

$$\left.\begin{aligned} &\hat{\sigma}_x\hat{\sigma}_y = -\hat{\sigma}_y\hat{\sigma}_x = i\hat{\sigma}_z, \quad \hat{\sigma}_y\hat{\sigma}_z = -\hat{\sigma}_z\hat{\sigma}_y = i\hat{\sigma}_x, \quad \hat{\sigma}_z\hat{\sigma}_x = -\hat{\sigma}_x\hat{\sigma}_z = i\hat{\sigma}_y; \\ &\hat{\sigma}_x^2 = \hat{\sigma}_y^2 = \hat{\sigma}_z^2 = \hat{1}\,; \end{aligned}\right\} \qquad \ldots\ldots(8.3b)$$

$$\mathrm{Tr}\,\hat{\sigma}_x\hat{\sigma}_y = \mathrm{Tr}\,\hat{\sigma}_y\hat{\sigma}_z = \mathrm{Tr}\,\hat{\sigma}_z\hat{\sigma}_x = 0, \quad \mathrm{Tr}\,\hat{\sigma}_x^2 = \mathrm{Tr}\,\hat{\sigma}_y^2 = \mathrm{Tr}\,\hat{\sigma}_z^2 = 2. \qquad \ldots\ldots(8.3c)$$

As $\mathbf{P}$ has three components we can use these as the three independent parameters to determine the density matrix $\hat{\rho}$. As $\hat{\rho}$ is a 2 by 2 matrix, we can express it in terms of the unit matrix $\hat{1}$ and the Pauli matrices,

$$\hat{\rho} = a.\hat{1} + (\mathbf{a}.\hat{\boldsymbol{\sigma}}). \qquad \ldots\ldots(8.4)$$

From the normalization condition (2.9) and Eqn (8.3*a*) we get

$$\mathrm{Tr}\,\hat{\rho} = 1 = 2a, \quad \text{or,} \quad a = \tfrac{1}{2}, \qquad \ldots\ldots(8.5)$$

while Eqn (8.1) leads with the aid of Eqns (8.3*a*) and (8.3*c*) to

$$\mathbf{P} = \langle\hat{\boldsymbol{\sigma}}\rangle = \mathrm{Tr}\,\hat{\rho}\hat{\boldsymbol{\sigma}} = \mathrm{Tr}\,[a\hat{\boldsymbol{\sigma}} + \hat{\boldsymbol{\sigma}}(\mathbf{a}.\hat{\boldsymbol{\sigma}})] = 2\mathbf{a}. \qquad \ldots\ldots(8.6)$$

Combining Eqns (8.4) to (8.6) we get

$$\hat{\rho} = \tfrac{1}{2}[\hat{1} + (\mathbf{P}.\hat{\boldsymbol{\sigma}})] = \tfrac{1}{2}\begin{Vmatrix} 1+P_z & P_x - iP_y \\ P_x + iP_y & 1-P_z \end{Vmatrix}. \qquad \ldots\ldots(8.7)$$

This equation shows that, indeed, $\hat{\rho}$ is determined, once $\mathbf{P}$ is known.

Let us now consider the case where this beam passes through a magnetic field, and ask what will happen to the polarization of the beam. We could use Eqn (2.37) for the rate of change of the density matrix and as $\hat{\rho}$ contains $\mathbf{P}$ obtain in that way the equation of motion for $\mathbf{P}$. The drawback of this procedure is that one cannot apply it with the same ease to the case of particles with spin greater than $\frac{1}{2}$. We shall instead use Eqn (2.38) for the rate of change of average values. From this equation and Eqn (8.11) we get

$$\frac{\partial\mathbf{P}}{\partial t} = \frac{\partial\langle\hat{\boldsymbol{\sigma}}\rangle}{\partial t} = -\frac{i}{\hbar}\langle[\hat{\boldsymbol{\sigma}}, \hat{H}_s]_-\rangle, \qquad \ldots\ldots(8.8)$$

where $\hat{H}_s$ is the Hamiltonian referring to the spin coordinate (the *spin Hamiltonian*) which is given by the equation

$$\hat{H}_s = -(\hat{\boldsymbol{\mu}}.\mathscr{H}) = -\tfrac{1}{2}\gamma\hbar(\hat{\boldsymbol{\sigma}}.\mathscr{H}), \qquad \ldots\ldots(8.9)$$

where $\mathscr{H}$ is the magnetic field, $\hat{\boldsymbol{\mu}}$ the magnetic moment of the electron ($\hat{\boldsymbol{\mu}} = e\hbar\hat{\boldsymbol{\sigma}}/2mc$ with e, electronic charge; m, electronic mass; c, velocity of light), and γ the magnetogyric ratio ($\gamma = e/mc$). If we write Eqns (8.3*b*) in the symbolical form

$$[\hat{\boldsymbol{\sigma}} \wedge \hat{\boldsymbol{\sigma}}] = 2i\hat{\boldsymbol{\sigma}}, \qquad \ldots\ldots(8.10)$$

we find from Eqns (8.8) and (8.9)

$$\partial \mathbf{P}/\partial t = \tfrac{1}{2}i\gamma\langle[\hat{\mathbf{\sigma}}, (\hat{\mathbf{\sigma}}.\mathscr{H})]_-\rangle$$
$$= \tfrac{1}{2}i\gamma\langle[\mathscr{H}\wedge(\hat{\mathbf{\sigma}}\wedge\hat{\mathbf{\sigma}})]\rangle = -\gamma[\mathscr{H}\wedge\langle\hat{\mathbf{\sigma}}\rangle],$$

or,
$$\partial \mathbf{P}/\partial t = -\gamma[\mathscr{H}\wedge\mathbf{P}], \qquad \ldots\ldots(8.11)$$

which is just the classical equation of motion for the polarization vector. One could prove Eqn (8.11) starting from the Schrödinger equation, but the proof is cumbersome. The ease with which we could prove Eqn (8.11) is an example of the advantages of the density matrix. Eqn (8.11) itself is a consequence of the generalized Ehrenfest theorem (Ehrenfest 1927, Kramers 1957, § 30) which states that any quantum-mechanical average will obey the corresponding classical equation of motion.

We can easily generalize the discussion leading to Eqn (8.11) to the case of larger spin values. Let $j\,(>\frac{1}{2})$ be the largest possible value of the angular momentum of the particles. The density matrix will now be a $2j+1$ by $2j+1$ matrix, and apart from the components of the polarization vector $\mathbf{P}$ we need other quantities to determine $\hat{\rho}$ completely. The vector $\mathbf{P}$ can be called the dipole polarization vector, and the other quantities which can be used to determine $\hat{\rho}$ are the quadrupole polarization tensor (5 components; its components and the 3 components of $\mathbf{P}$ are the 8 parameters which are sufficient to determine $\hat{\rho}$ if $j=1$), the octupole polarization tensor (7 components; it comes into play if $j>1$), ..., in general the 2^l-polarization tensor (with $2l+1$ components; for a given j, all multipoles with $l\leqslant 2j$ will be involved). We do not have the space here to go into a detailed discussion of the determination of the density matrix for this general case and refer to Fano's review article (1957) where further references can be found. The polarization vector $\mathbf{P}$ is now defined by the equation

$$\mathbf{P} = \langle\hat{\mathbf{J}}\rangle/j\hbar, \qquad \ldots\ldots(8.12)$$

where $\hat{\mathbf{J}}$ is the angular momentum. It satisfies the commutation relation which can be expressed by a formal equation similar to Eqn (8.10),

$$[\hat{\mathbf{J}}\wedge\hat{\mathbf{J}}] = i\hbar\hat{\mathbf{J}}. \qquad \ldots\ldots(8.13)$$

The spin Hamiltonian is given by the equation

$$\hat{H}_s = -\gamma(\hat{\mathbf{J}}.\mathscr{H}), \qquad \ldots\ldots(8.14)$$

and from Eqns (8.12), (8.13), (8.14) and (2.38) we find that Eqn (8.11) holds also for the general case.

Let us now consider the case of the polarization of electromagnetic radiation. In this case there are also two degrees of freedom, and we can thus expect a formal similarity between this case and the case of the electron beam. In both cases the wave functions describing the systems in the ensemble can be written in the form

$$\psi^k = c_1{}^k\psi_1 + c_2{}^k\psi_2, \quad k = 1, 2, \ldots, \qquad \ldots\ldots(8.15)$$

where ψ_1 and ψ_2 are two orthogonal wave functions. In the case of electrons ψ_1 may describe either the case where the electron spin is in the $+z$-direction, or the case where the electron spin is parallel to the electron momentum. The function ψ_2

will then describe the case where the electron spin is in the $-z$-direction, or the case where it is antiparallel to the electron momentum. In the case of photons ψ_1 could describe either a plane polarized wave or a right-hand circularly polarized wave, and ψ_2 would then describe a plane polarized wave with its plane of polarization perpendicular to the plane of polarization of the wave described by ψ_1, or a left-hand circularly polarized wave.

Each of the ψ^k corresponds to a totally polarized beam. If we were dealing with a pure state all ψ^k would be the same, say, equal to ψ_0:

$$\psi_0 = c_1^{(0)}\psi_1 + c_2^{(0)}\psi_2. \qquad \ldots\ldots(8.16)$$

In § 2 we mentioned that one can find a filter such that it would correspond exactly to a given pure case; for linearly polarized light a Nicol prism would be such a filter. We shall call such an instrument here a detector although strictly speaking a filter is, of course, not used to detect (compare the discussion in § 10) and we shall describe it by a density matrix $\hat{\rho}^{\text{det}}$. To find the form of this density matrix we first of all note that it should correspond to a pure state with wave function

$$\psi^{\text{det}} = c_1^{\text{det}}\psi_1 + c_2^{\text{det}}\psi_2. \qquad \ldots\ldots(8.17)$$

The response of this detector to a state with wave function (8.16) will be given by the overlap of ψ_0 and ψ^{det}, and the probability W for a response will be given by the expression

$$W = \left|\int \psi_0^* \psi^{\text{det}}\, d\tau\right|^2 = \left|c_1^{0*} c_1^{\text{det}} + c_2^{0} c_2^{\text{det}}\right|^2. \qquad \ldots\ldots(8.18)$$

If we are not dealing with a pure state, we must introduce for the beam the density matrix $\hat{\rho}$, and we shall require that W is now given by the expectation value of $\hat{\rho}^{\text{det}}$,

$$W = \langle\hat{\rho}^{\text{det}}\rangle = \operatorname{Tr}\hat{\rho}\hat{\rho}^{\text{det}}. \qquad \ldots\ldots(8.19)$$

One can easily check that one obtains for W expression (8.18) for the case of a pure state corresponding to all ψ^k being equal to ψ_0 if we take for $\hat{\rho}^{\text{det}}$ the density matrix which would correspond to a wave function ψ^{det}:

$$\hat{\rho}^{\text{det}} = \begin{Vmatrix} |c_1^{\text{det}}|^2 & c_1^{\text{det}} c_2^{\text{det}*} \\ c_1^{\text{det}*} c_2^{\text{det}} & |c_2^{\text{det}}|^2 \end{Vmatrix}. \qquad \ldots\ldots(8.20)$$

If the detector is a filter corresponding exactly to the state ψ_0, ψ^{det} should be equal to ψ_0, and W reaches its maximum value 1. If now $\hat{\rho}$ also corresponded to the pure state ψ_0 ($\hat{\rho} = \hat{\rho}_0$) we find also from Eqn (8.19) that $W = \langle\hat{\rho}^{\text{det}}\rangle = \operatorname{Tr}\hat{\rho}_0^2 = \operatorname{Tr}\hat{\rho}_0 = 1$, where we have used the condition (2.26) for a pure state.

We have here assumed a complete response of the detector for the pure state ψ^{det} and we shall therefore have a zero response for the pure state with the wave function which is orthogonal to ψ^{det}. We do not have space to enter here into a discussion of the case where the detector is not ideal, but responds less than completely to the most favourable polarization and still responds to the least favourable (opposite) polarization. We refer to Fano's review (1957) for a discussion of this point.

We want to express $\hat{\rho}^{\text{det}}$ in slightly different form. We mentioned earlier that a given wave function such as (8.17) corresponds to a well-defined polarization.

Let $\mathbf{P}^{\text{det}}$ be the corresponding polarization vector. By arguments similar to those leading to Eqn (8.7) we can then write for $\hat{\rho}^{\text{det}}$:

$$\hat{\rho}^{\text{det}} = \tfrac{1}{2}[\hat{1} + (\mathbf{P}^{\text{det}} . \hat{\boldsymbol{\sigma}})]. \qquad \ldots\ldots(8.21)$$

Using Eqns (8.21), (8.19) and (8.7) we get immediately an expression for the response of the detector to a beam with a polarization $\mathbf{P}$:

$$\begin{aligned} W = \operatorname{Tr} \hat{\rho}\hat{\rho}^{\text{det}} &= \operatorname{Tr} \tfrac{1}{2}[\hat{1} + (\mathbf{P}^{\text{det}} . \hat{\boldsymbol{\sigma}})]\,[\hat{1} + (\mathbf{P} . \hat{\boldsymbol{\sigma}})] \\ &= \tfrac{1}{2}[1 + (\mathbf{P}^{\text{det}} . \mathbf{P})], \qquad \ldots\ldots(8.22) \end{aligned}$$

where we have used Eqns (8.3). Eqn (8.22) leads to $W = 1$, if $\mathbf{P}^{\text{det}} \| \mathbf{P}$ (note that we have $\mathbf{P}$ defined in such a way that $\mathbf{P}^2 \leqslant 1$) and to $W = 0$, if $\mathbf{P}^{\text{det}} \| -\mathbf{P}$. In general, we can write

$$W = \tfrac{1}{2}(1 + P_n), \quad P_n = |\mathbf{P}| \cos\theta, \qquad \ldots\ldots(8.23)$$

where θ is the angle between $\mathbf{P}$ and $\mathbf{P}^{\text{det}}$.

The vector $\mathbf{P}$ has an immediate physical meaning in the case of electrons. In the case of photons, however, we must specify in more detail what is meant by $\mathbf{P}$ and also what is meant by ' opposite ' polarization, a term used a moment ago. So far we have only used the formal similarity between electron spin and photon polarization. In the case of photons $\mathbf{P}$ is a vector in a symbolical space (compare the isotopic spin vector in nuclear physics) which is introduced in order that we can continue to use the same formalism. The simplest way to introduce $\mathbf{P}$ for photons is to consider the complex vector potential $\mathbf{A} \exp[i(\mathbf{k}.\mathbf{x})]$ ($\mathbf{k}$ is the wave vector of the photon) as the wave function ψ. Instead of Eqn (8.16) we have now

$$\mathbf{A} \exp[i(\mathbf{k}.\mathbf{x})] = a_1 \mathbf{A}_1 \exp[i(\mathbf{k}.\mathbf{x})] + a_2 \mathbf{A}_2 \exp[i(\mathbf{k}.\mathbf{x})], \qquad \ldots\ldots(8.24)$$

where

$$(\mathbf{A}_1 . \mathbf{k}) = (\mathbf{A}_2 . \mathbf{k}) = (\mathbf{A}_1 . \mathbf{A}_2) = 0, \quad \mathbf{A}_1^2 = \mathbf{A}_2^2 = \mathbf{A}^2, \qquad \ldots\ldots(8.25)$$

which means that we have chosen ψ_1 and ψ_2 to describe two plane polarized waves with mutually perpendicular planes of polarization. The coefficients a_1 and a_2 completely describe the polarization of the photon. For this pure case, the density matrix is constructed by analogy with expression (8.20),

$$\rho = \begin{Vmatrix} |a_1|^2 & a_1 a_2^* \\ a_1^* a_2 & |a_2|^2 \end{Vmatrix} \qquad \ldots\ldots(8.26)$$

and $\mathbf{P}$ is defined by analogy with Eqn (8.7) :

$$P_1 = |a_1|^2 - |a_2|^2, \quad P_2 = a_1 a_2^* + a_1^* a_2, \quad P_3 = i(a_1 a_2^* - a_1^* a_2). \qquad \ldots\ldots(8.27)$$

To emphasize the fact that now $\mathbf{P}$ is defined in a ' polarization ' space rather than in physical space we have denoted its 3 components by P_1, P_2 and P_3, rather than by P_x, P_y and P_z as was done in the case of electrons. For a discussion of the relation between $\mathbf{P}$ and the so-called Stokes parameters (1852) or the coherence matrix used by Wolf and Roman we refer to the papers quoted at the beginning of this section.

We shall use the formalism developed here to discuss briefly the process where a positron and an electron in a singlet state annihilate one another with the simultaneous emission of two photons (compare the discussion of this process by Pryce and Ward (1947), Snyder, Pasternack and Hornbostel (1948) and Bleuler and

ter Haar (1948)). One can show that the two photons (A and B) which are emitted can be described by a joint density matrix $\hat{\rho}_{AB}$ (compare the discussion at the end of § 2) which is given by the equation

$$\hat{\rho}_{AB} = \tfrac{1}{4}[\hat{1}_A \hat{1}_B - (\hat{\boldsymbol{\omega}}_A . \hat{\boldsymbol{\omega}}_B)], \qquad \ldots\ldots(8.28)$$

where the subscripts A and B refer to the two photons, and the vectors $\hat{\boldsymbol{\omega}}$ are symbolical vectors whose components $\hat{\omega}_1$, $\hat{\omega}_2$ and $\hat{\omega}_3$ are two by two matrices which have the same form as the Pauli matrices (8.2) and which therefore satisfy equations similar to Eqns (8.3). We do not wish to give a proof of Eqn (8.28) but accepting its validity we want to investigate its physical meaning. To do this we assume that there are two detectors described by their density matrices $\hat{\rho}_A{}^{det}$ and $\hat{\rho}_B{}^{det}$ which can measure the polarization vectors $\mathbf{P}_A{}^{det}$ and $\mathbf{P}_B{}^{det}$. This means that (compare Eqn (8.21) and the definition of $\hat{\boldsymbol{\omega}}_A$ and $\hat{\boldsymbol{\omega}}_B$)

$$\hat{\rho}_A{}^{det} = \tfrac{1}{2}[\hat{1}_A + (\mathbf{P}_A{}^{det} . \hat{\boldsymbol{\omega}}_A)], \quad \hat{\rho}_B{}^{det} = \tfrac{1}{2}[\hat{1}_B + (\mathbf{P}_B{}^{det} . \hat{\boldsymbol{\omega}}_B)]. \qquad \ldots\ldots(8.29)$$

If we wish to know the polarization of photon A we must evaluate the expression

$$W_A = \mathrm{Tr}_{AB} \hat{\rho}_{AB} \hat{1}_A \hat{\rho}_B{}^{det} = \tfrac{1}{2} \qquad \ldots\ldots(8.30)$$

which corresponds to zero polarization. This we could also have seen by using Eqn (2.51) to obtain the density matrix $\hat{\rho}_A$,

$$\hat{\rho}_A = \mathrm{Tr}_B \hat{\rho}_{AB} = \tfrac{1}{2}\hat{1}_A, \qquad \ldots\ldots(8.31)$$

which in terms of ψ_1 and ψ_2 means that there is just as much polarization in one direction as in the opposite direction. As all our expressions are symmetric in A and B we find, of course, that the photon B is also unpolarized.

Let us now consider whether there is any correlation between the two polarizations. To do this we evaluate

$$W_{AB} = \mathrm{Tr}_{AB} \hat{\rho}_{AB} \hat{\rho}_A{}^{det} \hat{\rho}_B{}^{det} = \tfrac{1}{4}[1 - (\mathbf{P}_A{}^{det} . \mathbf{P}_B{}^{det})], \qquad \ldots\ldots(8.32)$$

which expresses the fact that the two photons have opposite polarization: W_{AB} is a maximum when $\mathbf{P}_A{}^{det} \| -\mathbf{P}_B{}^{det}$ and vanishes when $\mathbf{P}_A{}^{det} \| \mathbf{P}_B{}^{det}$.

To conclude this section we shall briefly discuss scattering and correlation experiments in the density matrix frame-work. Let us first of all consider the case where both the detector and the initial state of the system are pure states with wave functions ψ^{det} and $\psi(0)$. The scattering is described by the Schrödinger equation

$$i\hbar\dot{\psi}(t) = \hat{H}\psi(t) \qquad \ldots\ldots(8.33)$$

with the solution

$$\psi(t) = [\exp(-i\hat{H}t/\hbar)]\,\psi(0). \qquad \ldots\ldots(8.34)$$

The probability of detection is given by the equation

$$W(t) = \left| \int \psi^*(t)\, \psi^{det}\, d\tau \right|^2, \qquad \ldots\ldots(8.35)$$

or, expressed in the coefficients $c_i(t)$ and $c_i{}^{det}$ of an expansion of $\psi(t)$ and ψ^{det} in terms of some complete orthonormal set

$$W(t) = |\textstyle\sum_i c_i{}^{det}\, c_i{}^*(t)|^2 = \mathrm{Tr}\, \hat{\rho}^{det} \hat{\rho}(t), \qquad \ldots\ldots(8.36)$$

where we have used Eqn (2.6) for the definition of the elements of the density matrices. In a mixed state, Eqn (8.36) will still hold and $\hat{\rho}(t)$ will depend on $\hat{\rho}(0)$ through the solution of the equation of motion (2.37),

$$\hat{\rho}(t) = [\exp(-i\hat{H}t/\hbar)]\,\hat{\rho}(0)\exp(i\hat{H}t/\hbar). \qquad \ldots\ldots(8.37)$$

Let us now consider how we can obtain transition probabilities. In that case we are interested in detectors which are sensitive to systems within a small energy range of final states dE_t, and for the transition probability per unit time P, which is $W(t)/t$, we get from perturbation theory for the pure state case

$$P = (2\pi/\hbar)\left|\int \psi^{\text{det}}\hat{H}'\psi(0)\,d\tau\right|^2 dE_t, \qquad \ldots\ldots(8.38)$$

where we have assumed that we can use first-order perturbation theory, where $\hat{H}'$ is the perturbation Hamiltonian, and where we have assumed that the detector is not sensitive to unscattered particles.

If we are dealing with a mixed state we must generalize Eqn (8.38) and the result is easily seen to be

$$P = (2\pi/\hbar)\,dE_t\,\mathrm{Tr}\,[\hat{\rho}^{\text{det}}\hat{H}'\hat{\rho}\hat{H}'^{\dagger}], \qquad \ldots\ldots(8.39)$$

where $\hat{H}'^{\dagger}$ is the Hermitian conjugate of $\hat{H}'$.

Equation (8.39) is a general equation. The Hamiltonian $\hat{H}'$ is determined by the scattering process considered, $\hat{\rho}$ by the initial conditions, and $\hat{\rho}^{\text{det}}$ by the final conditions. If, for instance, we are interested in the angular correlation between a photon of a well-defined plane of polarization and an electron with a well-defined spin direction which are produced in a Compton scattering process, $\hat{\rho}^{\text{det}}$ will be the direct product of a factor of the form (8.21) for the electron, a factor of the form (8.29) for the photon, and a factor depending on the electron momentum and the photon wave vector. The actual evaluation of expression (8.39) may still be complicated (see, for instance, the papers by Lipps and Tolhoek (1954 a, b) on Compton scattering) but a great deal has been gained in not having to consider all degrees of freedom which are irrelevant to the final cross section in which we are interested.

§ 9. Resonance and Relaxation Phenomena

We saw in the preceding section that the density matrix is particularly suitable for discussing systems with a small number of degrees of freedom. It has, indeed, been very useful in discussions of various resonance phenomena involving a limited number of energy levels; recently, for instance, the theory of masers and maser-like devices has been discussed in density matrix terms (see, for instance, Anderson 1957, Clogston 1958, Suhl 1958). In the present section we shall restrict our discussion to the consideration of absorption of microwave radiation in a gas following Karplus and Schwinger's treatment (1948) and to a consideration of relaxation effects in nuclear magnetic induction following the classic paper by Wangsness and Bloch (1953). Among the steadily growing literature on the application of density matrix techniques to resonance and relaxation phenomena we may refer to papers by Karplus (1948), Bloch (1956, 1957), van Kranendonk

and Bloom (1956), Mori (1956), Seiden (1956 a, b), Wangsness (1956 a, b), Fano (1957), Kubo (1957 b), Redfield (1957, 1959), Abragam and Proctor (1958), Hubbard (1958), Kaplan (1958), Lamb and Sanders (1960), Lamb and Wilcox (1958), Schumacher (1958), Solomon (1958), Tomita (1958 a, b), Di Giacomo (1959), Kotera and Toda (1959), Skrotskiĭ and Kokin (1959), Wannier (1959), Yatsiv (1959), Halbach (1960), Nakano (1960 d), Sher and Primakoff (1960), Wilcox and Lamb (1960), Opechowski (1953) and Yvon (1960 a).

The first problem we shall consider is the evaluation of the absorption coefficient α for radiation of angular frequency ω. This absorption coefficient is proportional to the ratio of average dipole moment $\langle \hat{\boldsymbol{\beta}} \rangle$ to the field strength $\mathbf{F}(t)$. The field may be an electric or a magnetic field, and the total Hamiltonian of a particle in the gas will be of the form

$$\hat{H}(t) = \hat{H}_0 - (\hat{\boldsymbol{\beta}} . \mathbf{F}(t)) = \hat{H}_0 + \hat{V} \cos \omega t, \qquad \ldots\ldots(9.1)$$

where $\hat{\boldsymbol{\beta}}$ is the dipole moment operator, where $\hat{H}_0$ is the Hamiltonian of an isolated particle if no field is present, and where we have used the fact that

$$\mathbf{F}(t) = \mathbf{F} \cos \omega t. \qquad \ldots\ldots(9.2)$$

The state of the particles in the gas will be described by a density matrix $\hat{\rho}$ which satisfies the usual equation of motion (2.37) with the Hamiltonian given by Eqn (9.1). In Eqn (9.1) interactions between particles are not taken into account and we must therefore be careful to choose suitable boundary conditions for $\hat{\rho}$. We shall assume that collisions take place randomly with a mean time of flight τ, and that they are so strong that immediately after the collision $\hat{\rho}$ is given by the equilibrium expression (see Eqn (4.2)),

$$\hat{\rho}(t_0) = \hat{\rho}_0(t_0) = \exp[-\beta \hat{H}(t_0)] / \mathrm{Tr}\{\exp[-\beta \hat{H}(t_0)]\}, \qquad \ldots\ldots(9.3)$$

where t_0 is the time when the last collision took place. Between collisions the equation of motion for $\hat{\rho}$ will be (2.37),

$$i\hbar \dot{\hat{\rho}} = [\hat{H}, \hat{\rho}]_-, \qquad \ldots\ldots(9.4)$$

with $\hat{H}$ given by Eqn (9.1). In view of this splitting up of the changes in $\hat{\rho}$ into a smooth part governed by Eqn (9.4) and a discontinuous part which is governed by the boundary condition (9.3) it is useful to emphasize this fact and to write $\hat{\rho}$ as a function of two arguments $\hat{\rho}(t, t_0)$, where the second argument indicates when the last collision took place. The function $\hat{\rho}(t, t_0)$ thus satisfies Eqn (9.4) with the boundary condition

$$\hat{\rho}(t_0, t_0) = \hat{\rho}_0(t_0). \qquad \ldots\ldots(9.5)$$

The appearance of this second argument implies, however, that when we evaluate the average dipole moment we must take not only the usual average using the density matrix but also an average over all possible values of t_0. It turns out to be convenient to take this average first and to introduce a density matrix $\hat{\rho}_{\mathrm{Av}}$ defined by the equation

$$\hat{\rho}_{\mathrm{Av}}(t) = \int_0^\infty \hat{\rho}(t, t - \vartheta) \exp(-\vartheta/\tau)\, d\vartheta/\tau, \qquad \ldots\ldots(9.6)$$

where we have introduced the Poisson distribution $\exp(-\vartheta/\tau)\,d\vartheta/\tau$ for the probability that the last collision took place between $t-\vartheta$ and $t-\vartheta+d\vartheta$. Using $\hat{\rho}_{Av}$ we get for the average dipole moment

$$\langle\hat{\boldsymbol{\beta}}\rangle = \mathrm{Tr}\,\hat{\boldsymbol{\beta}}\hat{\rho}_{Av}. \qquad \ldots\ldots(9.7)$$

To evaluate $\langle\hat{\boldsymbol{\beta}}\rangle$ we need to know the equation of motion for $\hat{\rho}_{Av}$. As $\langle\hat{\boldsymbol{\beta}}\rangle$ vanishes when there is no field, it is convenient to introduce the difference $\widehat{\Delta\rho}$ between $\hat{\rho}_{Av}(t)$ and the density matrix $\hat{\rho}_0(t)$ for instantaneous thermal equilibrium. In terms of $\widehat{\Delta\rho}$, Eqn (9.7) becomes

$$\langle\hat{\boldsymbol{\beta}}\rangle = \mathrm{Tr}\,\hat{\boldsymbol{\beta}}\widehat{\Delta\rho}, \quad \widehat{\Delta\rho} = \hat{\rho}_{Av}(t) - \hat{\rho}_0(t). \qquad \ldots\ldots(9.8)$$

The equation of motion for $\widehat{\Delta\rho}$ can be found from Eqns (9.4) and (9.6),

$$\begin{aligned} i\hbar\widehat{\Delta\rho} &= -i\hbar\hat{\rho}_0 + i\hbar\hat{\rho}_{Av} \\ &= -i\hbar_0\hat{\rho} + \int_0^\infty i\hbar\left[\frac{\partial}{\partial t}\hat{\rho}(t,t_0)\right]_{t_0=t-\vartheta}\exp(-\vartheta/\tau)\frac{d\vartheta}{\tau} - \int_0^\infty i\hbar\frac{\partial}{\partial\vartheta}\hat{\rho}(t,t-\vartheta)\exp(-\vartheta/\tau)\frac{d\vartheta}{\tau} \\ &= -i\hbar\hat{\rho}_0 + [\hat{H},\widehat{\Delta\rho}]_- - \widehat{\Delta\rho}/\tau, \qquad \ldots\ldots(9.9) \end{aligned}$$

where we have evaluated the integral involving $(\partial/\partial\vartheta)$ by integrating by parts, used Eqn (9.6), and the fact that $[\hat{H},\hat{\rho}_0]_- = 0$. We must emphasize that this does not imply that $\hat{\rho}_0 = 0$, as $\hat{H}$ is an explicit function of the time.

We shall now assume that the field is weak so that we may neglect second-order terms in $\hat{V}$, $\widehat{\Delta\rho}$, or their product. This means that we can replace $[\hat{H},\widehat{\Delta\rho}]_-$ in Eqn (9.9) by $[\hat{H}_0,\widehat{\Delta\rho}]_-$. Using a representation in which $\hat{H}_0$ is diagonal, denoting the eigenvalues of $\hat{H}_0$ by E_n and introducing the notation

$$\hbar\omega_{mn} = E_m - E_n, \qquad \ldots\ldots(9.10)$$

we get from Eqn (9.9)

$$\left[\frac{\partial}{\partial t} + i\omega_{mn} + \frac{1}{\tau}\right]\Delta\rho_{mn} = -\frac{\partial}{\partial t}(\rho_0)_{mn}. \qquad \ldots\ldots(9.11)$$

If the field is weak, one can evaluate $(\rho_0)_{mn}$ using the Bloch equation (4.6)

$$\partial\hat{\mathfrak{p}}(\beta)/\partial\beta = (\hat{H}_0 + \hat{H}')\hat{\mathfrak{p}}(\beta), \qquad \ldots\ldots(9.12)$$

where $\hat{\mathfrak{p}}(\beta) = \exp(-\beta\hat{H})$ and $\hat{H}' \equiv \hat{V}\cos\omega t$. If we put

$$\hat{\mathfrak{p}}(\beta) = \hat{\mathfrak{p}}_0(\beta)\hat{S}(\beta), \quad \hat{\mathfrak{p}}_0(\beta) = \exp[-\beta\hat{H}_0], \qquad \ldots\ldots(9.13)$$

we get for $\hat{S}(\beta)$ the equation

$$\partial\hat{S}/\partial\beta = \hat{\mathfrak{p}}_0^{-1}\hat{H}'\hat{\mathfrak{p}}_0\hat{S}, \qquad \ldots\ldots(9.14)$$

which is equivalent to the integral equation

$$\hat{S}(\beta) = 1 + \int_0^\beta \hat{\mathfrak{p}}_0(\beta')^{-1}\hat{H}'\hat{\mathfrak{p}}_0(\beta')\,\hat{S}(\beta')\,d\beta', \qquad \ldots\ldots(9.15)$$

where we have used the fact that $\hat{\mathfrak{p}}(0) = \hat{\mathfrak{p}}_0(0) = 1$, or, $\hat{S}(0) = 1$. As $\hat{H}'$ is assumed to be small we can approximate Eqn (9.15) by

$$\hat{S}(\beta) \simeq 1 + \int_0^\beta \hat{\mathfrak{p}}_0(\beta')^{-1}\hat{H}'\hat{\mathfrak{p}}_0(\beta')\,d\beta'. \qquad \ldots\ldots(9.16)$$

Using Eqns (9.13) and the representation in which $\hat{H}_0$ is diagonal we get for $(\rho_0)_{mn}$

$$(\rho_0)_{mn} = \rho_m^{(0)}\delta_{mn} + (\rho_m^{(0)} - \rho_n^{(0)})(V_{mn}/\hbar\omega_{mn})\cos\omega t, \qquad \ldots\ldots(9.17)$$

where $\rho_m^{(0)}$ is given by the equation

$$\rho_m^{(0)} = \exp(-\beta E_m)/[\Sigma_n \exp(-\beta E_n)], \qquad \ldots\ldots(9.18)$$

and the V_{mn} are the matrix elements of the operator $\hat{V}$ in the same representation. The $\rho_m^{(0)}$ are the diagonal elements of the equilibrium density matrix. In Eqn (9.17) we have used the fact that up to terms linear in the field

$$\mathrm{Tr}\exp(-\beta\hat{H}) = \mathrm{Tr}\exp(-\beta\hat{H}_0).$$

From Eqns (9.8), (9.11), (9.17) and (9.18) we can find $\langle\hat{\mathbf{p}}\rangle$ and thus the absorption coefficient. If ω lies near a particular resonance frequency, ω_0, say ($\hbar\omega_0 = E_a - E_b$), and if the other ω_{mn} are such that $\omega_{mn} - \omega_0$ is always much greater than $1/\tau$, we find for the absorption coefficient per particle

$$\alpha = \frac{4\pi\omega^2}{3c}\beta|p_{ab}|^2\frac{\tau^{-1}}{(\omega-\omega_0)^2+\tau^{-2}}\frac{1-\exp(-\beta\hbar\omega_0)}{\beta\hbar\omega_0}\frac{\exp(-\beta E_b)}{\Sigma_n\exp(-\beta E_n)} \qquad \ldots\ldots(9.19)$$

which shows that $1/\tau$ will be the resonance width (we have assumed E_a and E_b to be non-degenerate). The derivation of Eqn (9.19) is straightforward, but tedious.

The other topic we want to discuss here is the derivation of the Bloch nuclear induction equation (Bloch 1946),

$$\frac{d\mathbf{P}}{dt} = \gamma[\mathbf{P}\wedge\mathscr{H}] - \frac{\mathbf{i}P_x}{\tau_2} - \frac{\mathbf{j}P_y}{\tau_2} - \frac{\mathbf{k}(P_z - P_0)}{\tau_1} \qquad \ldots\ldots(9.20)$$

where $\mathbf{P}(t)$ is the nuclear polarization due to an external magnetic field $\mathscr{H}(t)$, which consists of a strong, constant field $\mathscr{H}_0$ in the z-direction and a relatively weak field $\mathscr{H}_1$,

$$\mathscr{H} = \mathbf{k}\mathscr{H}_0 + \mathscr{H}_1(t). \qquad \ldots\ldots(9.21)$$

The vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are unit vectors in the x-, y-, and z-directions, γ is again the magnetogyric ratio, and $\mathbf{P}_0$ is the equilibrium polarization in the field $\mathscr{H}_0$. The quantities τ_1 and τ_2 are the longitudinal and transverse relaxation times. If there were no field $\mathscr{H}_1$, P_x and P_y would tend to zero with a time constant τ_2, and P_z would tend to P_0 with a characteristic time τ_1.

Actually, we shall use a simpler model than the one discussed by Wangsness and Bloch, and as a result we shall find $\tau_1 = \tau_2 (= \tau)$ so that we do not need to specify the direction of the constant field, and instead of Eqn (9.20) we shall find

$$\frac{d\mathbf{P}}{dt} = \gamma[\mathbf{P}\wedge\mathscr{H}] - \frac{\mathbf{P}-\mathbf{P}_0}{\tau}, \qquad \ldots\ldots(9.22)$$

where $\mathbf{P}_0$ is the equilibrium polarization in the field $\mathscr{H}_0$, while we write for $\mathscr{H}$ instead of Eqn (9.21) simply

$$\mathscr{H} = \mathscr{H}_0 + \mathscr{H}_1(t). \qquad \ldots\ldots(9.23)$$

Our first simplification consists in assuming that we are dealing with particles of spin $\frac{1}{2}$; this entails that a great many troublesome terms which occur in the general case vanish, and that $\mathbf{P}$ is given by Eqn (8.1),

$$\mathbf{P} = \langle\hat{\boldsymbol{\sigma}}\rangle. \qquad \ldots\ldots(9.24)$$

The system we are considering consists of the spins which are embedded in a lattice—the surroundings or bath. We shall assume the spins to be non-interacting—except through the bath of which they are part. In that case we can just consider a system consisting of a single spin in interaction with a bath, and obtain the final results for our macroscopic system by taking suitable statistical averages. The Hamiltonian of this system is of the following form

$$\hat{H} = \hat{H}_{\mathrm{M}} + \hat{H}_{\mathrm{B}} + \hat{H}_{\mathrm{int}}, \qquad \ldots\ldots(9.25)$$

where $\hat{H}_{\mathrm{M}}$ is the magnetic energy of the spins in the magnetic field,

$$\hat{H}_{\mathrm{M}} = \hat{H}_{\mathrm{M}_0} + \hat{H}_{\mathrm{M}_1}, \quad \hat{H}_{\mathrm{M}_0} = \tfrac{1}{2}\gamma\hbar(\hat{\boldsymbol{\sigma}}.\mathscr{H}_0), \quad \hat{H}_{\mathrm{M}_1} = \tfrac{1}{2}\gamma\hbar(\hat{\boldsymbol{\sigma}}.\mathscr{H}_1), \qquad \ldots\ldots(9.26)$$

$\hat{H}_{\mathrm{B}}$ is the Hamiltonian of the bath, and $\hat{H}_{\mathrm{int}}$ the interaction Hamiltonian describing the interaction between the spin and the bath. We shall assume that this interaction takes place through random magnetic fields $\mathscr{H}_{\mathrm{rand}}(t)$ produced by the motion of charges in the bath (see Redfield 1957) so that

$$\hat{H}_{\mathrm{int}} = -\tfrac{1}{2}\gamma\hbar(\hat{\boldsymbol{\sigma}}.\mathscr{H}_{\mathrm{rand}}), \qquad \ldots\ldots(9.27)$$

where the fact that $\mathscr{H}_{\mathrm{rand}}$ is random in both direction and magnitude is expressed by the relations

$$\int_0^\infty \mathscr{H}_{\mathrm{rand}}(t)\,dt = 0, \qquad \ldots\ldots(9.28)$$

$$\int_0^\infty \mathscr{H}_{\mathrm{rand}\,i}(t)\mathscr{H}_{\mathrm{rand}\,j}(t-t')\,dt' = \lambda_{ij}. \qquad \ldots\ldots(9.29)$$

Eqn (9.29) defines the time-correlation of the random field. We shall later on make the usual assumption that Eqns (9.28) and (9.29) still hold if we replace the upper limit of the integral by a ' sufficiently large ' time. Assumption (9.27) is responsible for the fact that $\tau_1 = \tau_2$ so that Eqn (9.20) reduces to Eqn (9.22).

It is convenient to use a representation in which $\hat{H}_{\mathrm{M}_0}$ and $\hat{H}_{\mathrm{B}}$ are diagonal. The density matrix will then have matrix elements $\langle mb|\hat{\rho}|m'b'\rangle$ where $m\ (=\pm\frac{1}{2})$ is the magnetic quantum number, and b a quantum number describing the state of the bath. We shall assume that the bath is at all times in thermal equilibrium. If that is the case we can write

$$\langle mb|\hat{\rho}|m'b'\rangle = \delta_{bb'}\cdot P(b)\langle m|\hat{\rho}_{\mathrm{sp}}|m'\rangle \qquad \ldots\ldots(9.30)$$

where $\hat{\rho}_{\mathrm{sp}}$ is a reduced spin density matrix,

$$\hat{\rho}_{\mathrm{sp}} = \mathrm{Tr}_b\,\hat{\rho}, \qquad \ldots\ldots(9.31)$$

Tr_b indicating the trace over the bath degrees of freedom, and $P(b)$ are the normalized Boltzmann factors $[\Sigma_b P(b) = 1]$.

We shall now consider $\hat{H}_{\mathrm{M}_0}$ and $\hat{H}_{\mathrm{B}}$ to be large compared with $\hat{H}_{\mathrm{M}_1}$ and $\hat{H}_{\mathrm{int}}$, and treat the latter as perturbations. It is further convenient to work in the interaction representation. Operators in the interaction representation will be indicated by an asterisk,

$$\hat{\Omega}^* = \hat{B}^{-1}\hat{A}^{-1}\hat{\Omega}\hat{A}\hat{B}, \qquad \ldots\ldots(9.32)$$

where

$$\hat{A} = \exp(-i\hat{H}_{\mathrm{M}_0}t/\hbar), \qquad \ldots\ldots(9.33)$$

$$\hat{B} = \exp(-i\hat{H}_{\mathrm{B}}t/\hbar). \qquad \ldots\ldots(9.34)$$

We note for future reference that $\hat{A}$ and $\hat{B}$ commute, and that if $\hat{\Omega}$ depends on $\hat{\sigma}$ only, and not on the bath degrees of freedom, Eqn (9.32) reduces to

$$\hat{\Omega}^* = \hat{A}^{-1}\hat{\Omega}\hat{A}. \qquad \ldots\ldots(9.35)$$

From the equation of motion for $\hat{\rho}$

$$i\hbar\partial\hat{\rho}/\partial t = [\hat{H}_{M_s} + \hat{H}_{M_i} + \hat{H}_B + \hat{H}_{int}, \hat{\rho}]_- \qquad \ldots\ldots(9.36)$$

we get for $\hat{\rho}^*$ the equation of motion

$$i\hbar\partial\hat{\rho}^*/\partial t = [\hat{H}_{M_i}{}^* + \hat{H}_{int}{}^*, \hat{\rho}^*]_-. \qquad \ldots\ldots(9.37)$$

From Eqn (9.37) we can find $\partial\hat{\rho}_{sp}{}^*/\partial t$ and then $\partial\hat{\rho}_{sp}/\partial t$, using Eqn (9.35), and so finally $\partial\mathbf{P}/\partial t$. We must draw attention to the fact that as $\hat{\rho}_{sp}$ is independent of the bath degrees of freedom, relation (9.31) also holds between $\hat{\rho}^*$ and $\hat{\rho}_{sp}{}^*$.

As the eigenvalues of $\hat{A}$ are $e^{-i\omega t}$, where

$$\omega = \gamma|\mathscr{H}_0|, \qquad \ldots\ldots(9.38)$$

we get from Eqn (9.35) with $\hat{\Omega} \equiv \hat{\rho}_{sp}$

$$\left\langle m\left|i\hbar\frac{\partial\hat{\rho}_{sp}}{\partial t}\right|m'\right\rangle = \langle m|[\hat{H}_{M_s}, \hat{\rho}_{sp}]_-|m'\rangle + \exp[i\omega(m-m')t]\left\langle m\left|i\hbar\frac{\partial\hat{\rho}_{sp}{}^*}{\partial t}\right|m'\right\rangle. \qquad \ldots\ldots(9.39)$$

To find $\partial\hat{\rho}_{sp}{}^*/\partial t$ we evaluate the change in $\hat{\rho}_{sp}{}^*$ during a ' small ' time interval Δt. This increment $\widehat{\Delta\rho_{sp}}{}^*$ is obtained from Eqn (9.37) by first evaluating $\widehat{\Delta\rho}^*$ and then taking the trace over the bath degrees of freedom. From Eqn (9.37) we find

$$\widehat{\Delta\rho}^* = \widehat{\Delta_1\rho}^* + \widehat{\Delta_2\rho}^*, \qquad \ldots\ldots(9.40)$$

where

$$i\hbar\widehat{\Delta_1\rho}^* = \int_0^{\Delta t}[\hat{H}_{M_i}{}^*(t), \hat{\rho}^*(0)]_- dt + \int_0^{\Delta t}[\hat{H}_{int}{}^*(t), \hat{\rho}^*(0)]_- dt \qquad \ldots\ldots(9.41)$$

is the first-order change and where $\widehat{\Delta_2\rho}^*$ is the second-order change in $\hat{\rho}^*$. To simplify Eqn (9.41) we assume that Δt though ' small ' is ' sufficiently large ', so that we can use Eqn (9.28). The second integral then vanishes and as $\hat{H}_{M_i}{}^*$ is independent of the bath degree of freedom we get from Eqn (9.41)

$$i\hbar\Delta_1\hat{\rho}_{sp}{}^* \simeq [\hat{H}_{M_i}{}^*, \hat{\rho}_{sp}{}^*]_-\Delta t. \qquad \ldots\ldots(9.42)$$

There is as yet no term depending on the bath, and we must go to the second order. In $\widehat{\Delta_2\rho}^*$ we can, however, neglect all terms involving $\hat{H}_{M_i}$ so that we are left with

$$\widehat{\Delta_2\rho}^* = -\hbar^{-2}\int_0^{\Delta t}dt'\int_0^{t'}[\hat{H}_{int}{}^*(t'), [\hat{H}_{int}{}^*(t''), \hat{\rho}^*(0)]_-]_- dt''. \qquad \ldots\ldots(9.43)$$

From this equation and Eqn (9.29) we get, provided Δt satisfies again certain conditions as to smallness and largeness,

$$i\hbar\Delta_2\hat{\rho}_{sp}{}^* \simeq \Sigma_{i,j}\lambda_{ij}'[\hat{\sigma}_i{}^*, [\hat{\sigma}_j{}^*, \hat{\rho}_{sp}{}^*]_-]_-\Delta t, \qquad \ldots\ldots(9.44)$$

where the λ_{ij}' are determined by the λ_{ij} of Eqn (9.29). Substituting Eqns (9.42) and (9.44) into Eqn (9.39) and putting $\partial\hat{\rho}_{sp}^*/\partial t = \widehat{\Delta\rho_{sp}}^*/\Delta t$, expression (9.42) combines with the first term on the right-hand side of Eqn (9.39) to give $\langle m|[\hat{H}_M, \hat{\rho}_{sp}]_-|m'\rangle$ and this term leads, as we saw in the preceding section, to the first term on the right-hand side of Eqn (9.22). The second term is more complicated. The general procedure is as follows: one uses Eqn (9.35) to go back from the starred to the unstarred operators and expands the double commutator in Eqn (9.44). After that one multiplies the result by $\hat{\boldsymbol{\sigma}}$ and evaluates the trace over the spin variables using the relations (8.3) between the components of $\hat{\boldsymbol{\sigma}}$. The result is equal to a constant times $[\langle\hat{\boldsymbol{\sigma}}\rangle - \mathbf{a}]$, where the constant can be identified with $-\tau^{-1}$ and is determined by the λ_{ij}, showing once again the general importance of correlation functions in statistical physics, and where $\mathbf{a}$ can be shown to be equal to $\mathbf{P}_0$. We refer to Wangsness and Bloch's paper for details of this evaluation.

§ 10. The Theory of Measurement

We shall finish this article with a brief discussion of the theory of measurement. This is of less practical importance than some of the applications with which we have been concerned in the preceding sections. It is, however, of theoretical importance and the present author felt that a survey article such as the present would be incomplete without this concluding section. A related topic is quantum-mechanical ergodic theory, and we shall have occasion briefly to mention that topic as well. The classic papers on the theory of measurement in quantum mechanics are those by von Neumann, summarized in his monograph (1932) and the beautiful exposé by London and Bauer (1939). Among recent contributions to the subject we may mention papers by Ludwig (1953, 1954), Baker (1958), Band (1958), Green (1958), Daneri and Loinger (unpublished) and Wakita (1960), while we refer to papers by Sakai (1937), Tolman (1938), ter Haar (1954, 1955), van Kampen (1954, 1956), Van Hove (1957, 1959), Bocchieri and Loinger (1959) and Caldirola (1959) for a discussion of quantum-mechanical ergodic theory.

A measuring process consists essentially of two separate parts—a fact which was pointed out by Margenau (1950) and which is not always sufficiently stressed. The first part is the preparation for the measurement which consists in the ideal case in preparing a pure case. This can be done—as we have discussed in earlier sections—by letting the system pass through a suitable filter. After that the system must still be observed to be in the particular pure state which has been prepared. To fix the ideas let us consider a beam of spin 1 particles, a case discussed by London and Bauer which we shall use to illustrate the ideas of the present section. If we wish to find out the polarization of such a beam, we can perform a Stern–Gerlach experiment and let the beam pass through a region in space in which there is a magnetic field. This field will split the original beam into three beams, each of which corresponds to a well-defined value of the spin component in the direction of the magnetic field. This splitting into three beams itself is not a measurement, but is the preparation for the measurement which will consist in having three counters, say, at well-defined positions in space, such that the clicking of one of them corresponds to one of the three possibilities for the spin component. Once

the actual measurement, or observation, has taken place the pure case has been destroyed through the interaction between detector (counter) and system.

Let us consider the process of the preparation for the measurement in somewhat more detail. The actual measurement we shall assume to be done by a counter. There are a number of problems connected with the actual action of detectors (see, for instance, Green 1958) but we shall assume that the actual detection is straightforward. As far as the preparation process is concerned, we shall assume that the system is in a pure state before this process begins. This restriction is not necessary and can be removed but it enables us to discuss more clearly various consequences of the measuring process. If the system is in a pure state, it can be described by a wave function $\psi_0(x)$ where x stands collectively for all coordinates describing the system. Let $\hat{F}$ be the physical quantity the value of which we wish to measure, and let u_k and f_k be, respectively, its eigenfunctions and eigenvalues. It is convenient to expand ψ_0 in terms of the u_k,

$$\psi_0(x) = \sum_k a_k u_k(x). \qquad \ldots\ldots(10.1)$$

Let $\hat{P}$ be the operator corresponding to the apparatus preparing the system for a measurement, and let $v_m(y)$ and p_m be, respectively, its eigenfunctions and eigenvalues, where y stands collectively for all coordinates describing the apparatus. In order that our apparatus should be suitable for our purpose we must assume that the eigenvalues p_m can be assigned to different channels such that each channel corresponds uniquely to a well-defined eigenvalue f_k. Before the measuring process takes place, the apparatus should be in a well-defined state, v_0 say, corresponding to a ' neutral ' situation. The combined wave function of system + apparatus will thus before the measuring process be of the form

$$\Psi_{\text{before}} = v_0(y) \sum_k a_k u_k(x). \qquad \ldots\ldots(10.2)$$

Once the measuring process starts, the apparatus and the system are coupled and the wave function will be of the general form

$$\Psi_{\text{coupled}} = \sum_{km} b_{km} u_k(x) v_m(y). \qquad \ldots\ldots(10.3)$$

In view of the fact that our apparatus was chosen such that each p_m can be assigned to a definite f_k we can simplify expression (10.3) and write

$$\Psi_{\text{coupled}} = \sum_k b_k u_k(x) v_k(y), \qquad \ldots\ldots(10.4)$$

indicating that each u_k will uniquely lead to a well-defined state v_k, or, that if the system were in the state u_k it would definitely lead to the state v_k of the apparatus. As, moreover, in the case of an ideal measuring set-up—such as, for instance, the Stern–Gerlach experiment described earlier—the apparatus should not change the probability amplitudes, we must require that

$$b_k = a_k, \qquad \ldots\ldots(10.5)$$

and the wave function after the preparation of the measurement will be of the form

$$\Psi_{\text{uncoupled}} = \sum_k a_k u_k(x) v_k(y). \qquad \ldots\ldots(10.6)$$

We are once again in a familiar situation. We are interested only in part of the degrees of freedom of the combined system, and each part is no longer described by a wave function but by a density matrix corresponding to a mixture. The quantity $|a_k|^2$ is now both the probability that the system was originally (and still is) in the state u_k (see Eqns (10.2) and (10.6)) and also the probability that the apparatus is in the state v_k leading to an observation from which we conclude to the eigenvalue p_k and thus to the eigenvalue f_k.

We note that the preparation for the measurement, that is, the coupling of the system to a filter, changes the situation from one corresponding to a pure state to one corresponding to a mixture (which, by the way, increases the entropy as it increases the lack of detailed knowledge). After the uncoupling the detectors which are uniquely assigned to the different channels will determine the probabilities for the different eigenstates of the apparatus to be realized and thus will determine the probability distribution corresponding to the physical quantity $\hat{F}$. We have no space here to go into related problems such as the fact that the experiment must be repeated in order that we can find the probability distribution.

Let us consider again the Stern–Gerlach experiment on a beam of spin 1 particles. If there were no magnetic field, the beam would pass on undeflected. The same would be true, if the spin of the particles were in the direction of the magnetic field, while spin components $+\hbar$ or $-\hbar$ along the direction of the magnetic field would produce deflections of the beam in opposite directions. Before the coupling we have (compare Eqns (10.1) and (10.2))

$$\psi_0 = a_{-1}u_{-1} + a_0 u_0 + a_1 u_1, \qquad \ldots\ldots(10.7)$$

$$\Psi_{\text{before}} = v_0[a_{-1}u_{-1} + a_0 u_0 + a_1 u_1], \qquad \ldots\ldots(10.8)$$

where u_{-1}, u_0 and u_1 are the eigenfunctions of the spin component in the direction of the magnetic field with eigenvalues $-\hbar$, 0 and $\hbar$ respectively, while v_0 indicates that the beam would pass undeflected through if there were no spin–field interaction. Eqn (10.8) corresponds to Eqn (10.3) with

$$b_{km} = a_k \delta_{m0}, \qquad \ldots\ldots(10.9)$$

and to the pure state density matrices

$$(\rho_{\text{syst}})_{kl} = a_k^* a_l, \quad (\rho_{\text{app}})_{mn} = \delta_m \delta_{n0}. \qquad \ldots\ldots(10.10)$$

After the coupling and uncoupling we find (see Eqn (10.6))

$$\Psi_{\text{uncoupled}} = a_{-1}u_{-1}v_{-1} + a_0 u_0 v_0 + a_1 u_1 v_1 \qquad \ldots\ldots(10.11)$$

corresponding to

$$b_{km} = a_k \delta_{km}, \qquad \ldots\ldots(10.12)$$

and

$$(\rho_{\text{syst}})_{kl} = |a_k|^2 \delta_{kl}, \quad (\rho_{\text{app}})_{mn} = |a_m|^2 \delta_{mn}, \qquad \ldots\ldots(10.13)$$

as should be the case.

It is of interest to verify that the transition from (10.8) to (10.11) occurs through a unitary transformation, as should be the case, since it should be the evolution of the combined wave function under the influence of a Hamiltonian $\hat{H}$ given by the expression

$$\hat{H} = \hat{H}_{\text{syst}} + \hat{H}_{\text{app}} + \hat{H}_{\text{int}}, \qquad \ldots\ldots(10.14)$$

where $\hat{H}_{syst}$ and $\hat{H}_{app}$ are the unperturbed Hamiltonians of the system and of the apparatus, while $\hat{H}_{int}$ is an interaction term describing the coupling between system and apparatus. One can easily verify that

$$(b_{km})_{\text{uncoupled}} = \sum_{l,n} S_{km,ln} (b_{ln})_{\text{before}}, \qquad \ldots\ldots(10.15)$$

where $S_{km,ln}$ is the matrix

k, m \ l, n	$-1,-1$	$-1,0$	$-1,1$	$0,-1$	$0,0$	$0,1$	$1,-1$	$1,0$	$1,1$
$-1,-1$	0	1	0	0	0	0	0	0	0
$-1,0$	0	0	1	0	0	0	0	0	0
$-1,1$	1	0	0	0	0	0	0	0	0
$0,-1$	0	0	0	1	0	0	0	0	0
$0,0$	0	0	0	0	1	0	0	0	0
$0,1$	0	0	0	0	0	1	0	0	0
$1,-1$	0	0	0	0	0	0	0	0	1
$1,0$	0	0	0	0	0	0	1	0	0
$1,1$	0	0	0	0	0	0	0	1	0

Each of the three beams produced by the magnetic field corresponds to a well-defined value of the spin component, and thus to a well-defined pure case, but the total of the three beams corresponds to a mixture, produced by the magnetic field from the original pure case.

Concluding Remarks and Acknowledgments

In this paper I have given a bird's eye view of a few of the many aspects of density matrix theory. Lack of space has made it necessary to give only a sketch of many of the topics which were discussed. I hope to discuss this subject in more detail elsewhere.

I should like to express my gratitude to the great number of people who have kindly sent me preprints of relevant papers by themselves and their collaborators, and especially to Drs. Argyres, Brush, Golden, Kanazawe, Löwdin, McWeeny, Rosenfeld, Siegert, Thouless, Wolf and Yvon.

I should also like to express my thanks to Dr. W. E. Parry for helpful comments.

References

The number in brackets refer to those sections in this paper where the paper in question is referred to.

Abragam, A., and Proctor, W. G., 1958, *Phys. Rev.*, **109**, 1441 (9).

Abrikosov, A. A., Gor'kov, L. P., and Dzyaloshinskiĭ, I. E., 1959, *J. Exp. Theor. Phys., U.S.S.R.*, **36**, 900 (*Soviet Phys., JETP*, **9**, 636) (1, 3).

Adams, E. N., and Holstein, T. D., 1959, *J. Phys. Chem. Solids*, **10**, 254 (7).

Alfred, L. C. R., 1960, *Techn. Rep., No.* 7, *Project NR*-051–362, Brandeis University, U.S.A. (5).

Anderson, P. W., 1957, *J. Appl. Phys.*, **28**, 1049 (9).

Argyres, P. N., 1958, *Phys. Rev.*, **109**, 1115 (7); 1960 a, *Phys. Rev.*, **117**, 315 (7); 1960 b, *Westinghouse Sci. Paper* 6-94760-2-P16 (7).

ARGYRES, P. N., and ROTH, L. M., 1959, *J. Phys. Chem. Solids*, **12**, 89 (7).
AYRES, R. U., 1958, *Phys. Rev.*, **111**, 1453 (5).
BAKER, G. A., 1958, *Phys. Rev.*, **109**, 2198 (10).
BALESCU, R., 1959, *Physica*, **25**, 324 (7).
BALIAN, R., BLOCH, C., and DE DOMINICIS, C., 1960, *C.R. Acad. Sci., Paris*, **250**, 2850 (4).
BALIAN, R., and DE DOMINICIS, C., 1960 a, *Nucl. Phys.*, **16**, 502 (3, 4); 1960 b, *C.R. Acad. Sci., Paris*, **250**, 3285 (4); 1960 c, *C.R. Acad. Sci., Paris*, **250**, 4111 (4).
BAND, W., 1958, *Amer. J. Phys.*, **26**, 440 (10).
BELYAEV, S. T., 1958 a, b, *J. Exp. Theor. Phys., U.S.S.R.*, **34**, 417, 433 (*Soviet Phys., JETP*, **7**, 289, 299) (3).
BETHE, H. A., 1956, *Phys. Rev.*, **103**, 1353 (4).
BLATT, J. M., 1956, *Nuovo Cim.*, **4**, 430 (5).
BLATT, J. M., and MATSUBARA, T., 1958, *Progr. Theor. Phys. Japan*, **20**, 553 (4, 5).
BLEULER, E., and TER HAAR, D., 1948, *Science*, **108**, 10 (8).
BLOCH, C., and DE DOMINICIS, C., 1958, *Nucl. Phys.*, **7**, 459 (4); 1959 a, b, *Nucl. Phys.*, **10**, 181, 509 (4).
BLOCH, F., 1932, *Z. Phys.*, **74**, 295 (4); 1946, *Phys. Rev.*, **70**, 460 (9); 1956, *Phys. Rev.*, **102**, 104 (9); 1957, *Phys. Rev.*, **105**, 1206 (9).
BOCCHIERI, P., and LOINGER, A., 1959, *Phys. Rev.*, **114**, 948 (10).
BOGOLYUBOV, N. N., and SHIRKOV, D. V., 1959, *Introduction to the Theory of Quantized Fields* (New York: Interscience) (3, 4).
BOGOLYUBOV, N. N., and TYABLIKOV, S. V., 1959, *Dokl. Akad. Nauk, S.S.S.R.*, **126**, 53 (*Soviet Phys., Doklady*, **4**, 604) (3, 6).
BONCH-BRUEVICH, V. L., 1955, *J. Exp. Theor. Phys., U.S.S.R.*, **28**, 121 (*Soviet Phys., JETP*, **1**, 169) (3); 1956 a, *J. Exp. Theor. Phys., U.S.S.R.*, **31**, 522 (*Soviet Phys., JETP*, **4**, 457) (1, 3); 1956 b, *J. Exp. Theor. Phys., U.S.S.R.*, **30**, 342 (*Soviet Phys., JETP*, **3**, 278) (3); 1956 c, *J. Exp. Theor. Phys., U.S.S.R.*, **31**, 254 (*Soviet Phys., JETP*, **4**, 196) (3, 6); 1957, *Fiz. Met. Metallov.*, **4**, 546 (*Phys. Met. Metallogr.*, **4**, No. 3, 133) (3); 1958 a, b, *Fiz. Met. Metallov.*, **6**, 590, 769 (*Phys. Met. Metallogr.*, **6**, No. 4, 13 and No. 5, 1) (3); 1959 a, *Dokl. Akad. Nauk S.S.S.R.*, **124**, 1233 (*Soviet Phys., Doklady*, **4**, 172) (3); 1959 b, *Dokl. Akad. Nauk S.S.S.R.*, **126**, 539 (*Soviet Phys., Doklady*, **4**, 596) (3); 1959 c, *Dokl. Akad. Nauk S.S.S.R.*, **129**, 529 (*Soviet Phys., Doklady*, **4**, 1275) (3); 1959 d, *Fiz. Met. Metallov.*, **7**, 174 (*Phys. Met. Metallogr.*, **7**, No. 2, 13) (3); 1959 e, *J. Exp. Theor. Phys., U.S.S.R.*, **36**, 924 (*Soviet Phys., JETP*, **9**, 653) (7).
BONCH-BRUEVICH, V. L., and GLASKO, V. B., 1959, *Dokl. Akad. Nauk S.S.S.R.*, **124**, 1015 (*Soviet Phys., Doklady*, **4**, 147) (3).
BONCH-BRUEVICH, V. L., and KOGAN, SH. M., 1960, *Ann. Phys., New York*, **9**, 125 (3).
BONCH-BRUEVICH, V. L., and MIRONOV, A. G., 1960, *Fiz. Tverd. Tel.*, **2**, 489 (*Soviet Phys., Solid St.*, **2**, 454) (3).
BOPP, F., 1959, *Z. Phys.*, **156**, 348 (1, 5).
BORN, M., and GREEN, H. S., 1946, *Proc. Roy. Soc.* A, **188**, 10 (7); 1947 a, *Proc. Roy. Soc.* A, **190**, 455 (7); 1947 b, *Proc. Roy. Soc.* A, **191**, 168 (7); 1948, *Proc. Roy. Soc.* A, **192**, 166 (7).
BROUT, R., 1959, *Phys. Rev.*, **115**, 824 (4, 6).
BRUSH, S. G., 1961, *Rev. Mod. Phys.*, **33**, 79 (4).
BUTLER, S. T., and FRIEDMAN, M. H., 1955, *Phys. Rev.*, **98**, 287 (4).
CALDIROLA, P., 1959, *Nuovo Cim.*, **14**, 260 (10).
CALLEN, H. B., and WELTON, T. A., 1951, *Phys. Rev.*, **83**, 34 (7).
CARSLAW, H. S., and JAEGER, J. C., 1947, *Conduction of Heat in Solids* (Oxford: University Press), p. 217 (6).
CHEN' CHUN'-SYAN', 1959, *Dokl. Akad. Nauk S.S.S.R.*, **125**, 1238 (*Soviet Phys., Doklady*, **4**, 413) (3).
CHESTER, G. V., 1954, *Phys. Rev.*, **93**, 606 (4).
CHESTER, G. V., and THELLUNG, A., 1959, *Proc. Phys. Soc.*, **73**, 745 (1, 7).
CHIRGWIN, B. H., 1957, *Phys. Rev.*, **107**, 1013 (5).
CLOGSTON, A. M., 1958, *J. Phys. Chem. Solids*, **4**, 271 (9).

Corson, E. M., 1951, *Perturbation Methods in the Quantum Mechanics of n-electron Systems* (London : Blackie) (5).

Cox, J. A. M., and Tolhoek, H. A., 1953, *Physica*, **19**, 673 (8).

Di Giacomo, A., 1959, *Nuovo Cim.*, **14**, 1082 (9).

Dirac, P. A. M., 1929, *Proc. Camb. Phil. Soc.*, **25**, 62 (1, 5) ; 1930 a, b, *Proc. Camb. Phil. Soc.*, **26**, 361, 376 (1, 5) ; 1931, *Proc. Camb. Phil. Soc.*, **27**, 240 (1, 5) ; 1935, *Principles of Quantum Mechanics* (Oxford : University Press), 2nd edition, § 37 (1).

Dzyub, I. P., 1960, *Dokl. Akad. Nauk S.S.S.R.*, **130**, 1241 (*Soviet Phys., Doklady*, **5**, 125) (3).

Edwards, S. F., 1958 a, *Phil. Mag.*, **3**, 119 (6) ; 1958 b, *Phil. Mag.*, **3**, 1020 (1, 7).

Ehrenfest, P., 1927, *Z. Phys.*, **45**, 455 (8).

Ehrenreich, H., and Cohen, M. H., 1959, *Phys. Rev.*, **115**, 786 (5, 6).

Enz, C. P., 1960 a, b, *Helv. Phys. Acta*, **33**, 89, 115 (6).

Ezawa, H., Tomozawa, J., and Umezawa, H., 1957, *Nuovo Cim.*, **5**, 810 (3).

Falk, D. S., 1959 a, b, *Phys. Rev.*, **115**, 1069, 1074 (3).

Fano, U., 1949, *J. Opt. Soc. Amer.*, **39**, 859 (8) ; 1957, *Rev. Mod. Phys.*, **29**, 74 (1, 2, 8, 9).

Feynman, R. P., 1948, *Rev. Mod. Phys.*, **20**, 367 (4) ; 1953, *Phys. Rev.*, **91**, 1291 (4).

Fradkin, E. S., 1959 a, b, *J. Exp. Theor. Phys., U.S.S.R.*, **36**, 951, 1286 (*Soviet Phys., JETP*, **9**, 672, 912) (1, 3) ; 1959 c, *Nucl. Phys.*, **12**, 465 (1, 3, 5) ; 1959 d, e, *Dokl. Akad. Nauk S.S.S.R.*, **125**, 66, 311 (*Soviet Phys., Doklady*, **4**, 327, 347) (3).

Friedman, H. M., and Butler, S. T., 1955, *Phys. Rev.*, **98**, 294 (4).

Fujita, S., 1959, *Phys. Rev.*, **115**, 1335 (3, 4).

Fujita, S., and Hirota, R., 1960, *Phys. Rev.*, **118**, 6 (3, 4).

Galitskiĭ, V. M., 1958, *J. Exp. Theor. Phys., U.S.S.R.*, **34**, 151 (*Soviet Phys., JETP*, **7**, 104) (3).

Galitskiĭ, V. M., and Migdal, A. B., 1958, *J. Exp. Theor. Phys., U.S.S.R.*, **34**, 139 (*Soviet Phys., JETP*, **7**, 96) (3).

Gaudin, M., 1960 a, *Nucl. Phys.*, **15**, 89 (4) ; 1960 b, *Nucl. Phys.*, **20**, 513 (4).

Glassgold, A. E., Heckrotte, W., and Watson, K. M., 1959, *Phys. Rev.*, **115**, 1374 (4).

Goldberger, M. L., and Adams, E. N., 1952, *J. Chem. Phys.*, **20**, 240 (4).

Golden, S., 1957 a, *Phys. Rev.*, **105**, 604 (5) ; 1957 b, *Phys. Rev.*, **107**, 1283 (5) ; 1960 a, *Rev. Mod. Phys.*, **32**, 322 (1, 5) ; 1960 b, *Techn. Rep. No.* 6, *Project NR-051-362*, Brandeis University, U.S.A. (5).

Goldstone, J., 1957, *Proc. Roy. Soc.* A, **239**, 367 (4).

Goldstone, J., and Gottfried, K., 1959, *Nuovo Cim.*, **13**, 849 (5, 6).

Gombás, P., 1956, *Handbuch der Physik*, **36**, 109 (Berlin : Springer) (5).

Gor'kov, L. P., 1958, *J. Exp. Theor. Phys., U.S.S.R.*, **34**, 735 (*Soviet Phys., JETP*, **7**, 505) (3).

Green, H. S., 1947, *Proc. Roy. Soc.* A, **189**, 103 (7) ; 1958, *Nuovo Cim.*, **9**, 880 (1, 10).

Green, M. S., 1960, *Phys. Rev.*, **119**, 829 (7).

Greenwood, D. A., 1958, *Proc. Phys. Soc.*, **71**, 585 (1, 7).

de Groot, S. R., and Tolhoek, H. A., 1955, *Beta- and Gamma-Ray Spectroscopy* (Amsterdam : North-Holland), p. 613 (8).

Gurzhi, R. N., 1957, *J. Exp. Theor. Phys., U.S.S.R.*, **33**, 451 (*Soviet Phys., JETP*, **6**, 352) (7).

ter Haar, D., 1954, *Elements of Statistical Mechanics* (New York : Rinehart) (1, 2, 4, 6, 10) ; 1955, *Rev. Mod. Phys.*, **27**, 289 (1, 4, 10) ; 1958, *Introduction to the Physics of Many-Body Systems* (New York : Interscience) (5) ; 1960, *Physica*, **26**, 1041 (1, 5).

Hagedorn, R., 1958, *CERN Report No.* 58–7 (2, 8).

Halbach, K., 1960, *Phys. Rev.*, **119**, 1230 (9).

Hall, G. G., 1960, *Proc. Phys. Soc.*, **75**, 575 (5).

Hamilton, J., 1959, *The Theory of Elementary Particles* (Oxford : University Press) (3, 4, 5).

Hartogh, C. D., Tolhoek, H. A., and de Groot, S. R., 1954, *Physica*, **20**, 1310 (8).

Hebborn, J. E., and Sondheimer, E. H., 1960, *J. Phys. Chem. Solids*, **13**, 105 (6).

Helfand, E., 1960, *Phys. Rev.*, **119**, 1 (7).

Helfand, E., and Rice, S. A., 1960, *J. Chem. Phys.*, **32**, 1642 (7).

Hubbard, J., 1959, *Phys. Rev. Letters*, **3**, 77 (4, 5).

Hubbard, P. S., 1958, *Phys. Rev.*, **109**, 1153 (9).

Huby, R., 1958, *Proc. Phys. Soc.*, **72**, 97 (8).

HUGENHOLTZ, N. M., 1957, *Physica*, **23**, 481 (4).
HUGENHOLTZ, N. M., and PINES, D., 1959, *Phys. Rev.*, **116**, 489 (3).
HUND, F., 1956, *Handbuch der Physik*, **36**, 1 (Berlin : Springer) (5).
HUSIMI, K., 1940, *Proc. Phys. Math. Soc., Japan*, **22**, 264 (1, 2, 4, 5).
HUSIMI, K., KITANO, Y., and NISHIYAMA, T., 1958, *Fortschr. Phys.*, **6**, 1 (4, 5, 7).
ICHIMURA, H., 1956, *Progr. Theor. Phys., Japan*, **15**, 151 (6).
VAN KAMPEN, N. G., 1954, *Physica*, **20**, 603 (10) ; 1956, *Fortschr. Phys.*, **4**, 405 (7, 10).
KANAZAWA, H., and MATSUDAIRA, N., 1960, *Prog. Theor. Phys., Japan*, **23**, 433 (6).
KANAZAWA, H., MISAWA, S., and FUJITA, E., 1960, *Prog. Theor. Phys., Japan*, **23**, 426 (6).
KANAZAWA, H., and WATABE, M., 1959, *Progr. Theor. Phys., Japan*, **22**, 466 (3) ; 1960, *Progr. Theor. Phys., Japan*, **23**, 408 (6, 7).
KAPLAN, J. I., 1958, *J. Chem. Phys.*, **29**, 462 (9).
KARPLUS, R., 1948, *Phys. Rev.*, **73**, 1027 (9).
KARPLUS, R., and SCHWINGER, J., 1948, *Phys. Rev.*, **73**, 1020 (9).
KASCHLUHN, F., 1955 a, *Ann. Phys., Lpz.*, **16**, 257 (4) ; 1955 b, *Ann. Phys., Lpz.*, **16**, 304 (5).
KASUYA, T., 1959, *J. Phys. Soc., Japan*, **14**, 410 (7).
KEMBLE, E. C., 1937, *Fundamental Principles of Quantum Mechanics* (New York : McGraw-Hill) (2).
KINOSHITA, T., and NAMBU, Y., 1954, *Phys. Rev.*, **94**, 598 (3, 5).
KITTEL, C., 1958, *Elementary Statistical Physics* (New York : Wiley) (1).
KLEIN, A., and PRANGE, R., 1958, *Phys. Rev.*, **112**, 994 (3, 4).
KLEIN, A., and ZEMACH, C., 1957, *Phys. Rev.*, **108**, 126 (3).
KLIMONTOVICH, YU. L., and SILIN, V. P., 1960, *Usp. Fiz. Nauk*, **70**, 247 (*Soviet Phys., Uspekhi*, **3**, 84) (7).
KLIMONTOVICH, YU. L., and TEMKO, S. V., 1958, *J. Exp. Theor. Phys., U.S.S.R.*, **35**, 1141 (*Soviet Phys., JETP*, **8**, 799) (7).
KLINGER, M. I., 1959 a, *Fiz. Tverd. Tel.*, **1**, 674 (*Soviet Phys., Solid St.*, **1**, 613) (3, 7) ; 1959 b, c, d, *Fiz. Tverd. Tel.*, **1**, 861, 1226, 1385 (*Soviet Phys., Solid St.*, **1**, 782, 1122, 1269) (7) ; 1959 e, *Fiz. Tverd. Tel.*, 2nd collection of papers, p. 136 (7).
KOBELEV, L. YA., 1958 a, *Izv. Vyssh. Uch. Zaved.*, No. 1, 68 (3, 5) ; 1958 b, *Izv. Vyssh. Uch. Zaved.*, No. 2, 103 (3) ; 1958 c, d, *Fiz. Met. Metallov.*, **6**, 354, 750 (*Phys. Met. Metallogr.*, **6**, No. 2, 156 and No. 4, 165) (3) ; 1958 e, *Fiz. Met. Metallov.*, **6**, 943 (*Phys. Met. Metallogr.*, **6**, No. 5, 172) (3, 5) ; 1959, *Izv. Vyssh. Uch. Zaved.*, No. 2, 60 (3).
KOGAN, SH. M., 1959, *Dokl. Akad. Nauk, S.S.S.R.*, **126**, 546 (*Soviet Phys., Doklady*, **4**, 604) (3).
KOHN, W., and LUTTINGER, J. M., 1957, *Phys. Rev.*, **108**, 590 (1, 7).
KONSTANTINOV, O. V., and PEREL', V. I., 1959, *J. Exp. Theor. Phys., U.S.S.R.*, **37**, 786 (*Soviet Phys., JETP*, **10**, 560) (7) ; 1960, *J. Exp. Theor. Phys., U.S.S.R.*, **39**, 197 (*Soviet Phys., JETP*, **12**, 142) (7).
KOPPE, H., 1951, *Ann. Phys., Lpz.*, **9**, 423 (3, 4) ; 1957, *Z. Phys.*, **148**, 135 (5, 6).
KOPPE, H., and MÜHLSCHLEGEL, B., 1958, *Z. Phys.*, **151**, 613 (6).
KOTANI, T., 1955, *Prog. Theor. Phys., Japan*, **14**, 379 (4, 8).
KOTERA, T., and KODA, M., 1959, *J. Phys. Soc., Japan*, **14**, 1475 (9).
KRAICHNAN, R. H., 1958 a, b, *Phys. Rev.*, **112**, 1054, 1056 (3).
KRAMERS, H. A., 1957, *Quantum Mechanics* (Amsterdam : North-Holland) (5, 8).
VAN KRANENDONK, J., and BLOOM, M., 1956, *Physica*, **22**, 545 (9).
KRIVOGLAZ, M. A., and PEKAR, S. I., 1957 a, b, *Izv. Akad. Nauk, S.S.S.R., Ser. Fiz.*, **21**, 3, 16 (6).
KUBO, R., 1952, *J. Chem. Phys.*, **20**, 770 (4) ; 1956, *Canad. J. Phys.*, **34**, 1274 (7) ; 1957 a, *J. Phys. Soc., Japan*, **12**, 570 (1, 3, 7) ; 1957 b, *Nuovo Cim. Suppl.*, **6**, 1063 (9).
KUBO, R., HASEGAWA, H., and HASHITSUME, N., 1959, *J. Phys. Soc., Japan*, **14**, 56 (7).
KÜMMEL, H., 1955, *Z. Phys.*, **143**, 219 (7).
LAMB, W. E., and SANDERS, T. M., 1960, *Phys. Rev.*, **119**, 1901 (9).
LAMB, W. E., and WILCOX, L., 1958, *J. Phys. Radium*, **19**, 801 (9).
LANDAU, L. D., 1930, *Z. Phys.*, **64**, 629 (6) ; 1958, *J. Exp. Theor. Phys.*, **34**, 262 (*Soviet Phys., JETP*, **7**, 182) (3).

LANDAU, L. D., and LIFSHITZ, E. M., 1958 a, *Quantum Mechanics* (London: Pergamon) (1); 1958 b, *Statistical Physics* (London: Pergamon) (1).
LAX, M., 1958, *Phys. Rev.*, **109**, 1921 (1, 7).
LEE, T. D., and YANG, C. N., 1959, *Phys. Rev.*, **113**, 1165 (4).
VAN LEEUWEN, J. M. J., GROENEVELD, J., and DE BOER, J., 1959, *Physica*, **25**, 792 (4).
LEVINE, H. B., 1960, *Phys. Fluids*, **3**, 225 (4).
LIPPS, F. W., and TOLHOEK, H. A., 1954 a, b, *Physica*, **20**, 85, 395 (8).
LONDON, F., and BAUER, E., 1939, *La Théorie de l'Observation en Méchanique Quantique* (Paris: Hermann) (1, 2, 10).
LÖWDIN, P.-O., 1954, *Proc. 10th Solvay Congress* (5); 1955 a, b, c, *Phys. Rev.*, **97**, 1474, 1490, 1509 (5); 1956, *Advanc. Phys.*, **5**, 1 (5); 1959, *Advanc. Chem. Phys.*, **2**, 207 (5); 1960, *Rev. Mod. Phys.*, **32**, 328 (1, 5).
LUDWIG, G., 1953, *Z. Phys.*, **135**, 483 (10); 1954, *Grundlagen der Quantenmechanik* (Berlin: Springer) (1, 10).
LUNDQUIST, S. O., 1959, *Ark. Fys.*, **16**, No. 30 (4).
LUTTINGER, J. M., 1958, *Phys. Rev.*, **112**, 739 (7).
LUTTINGER, J. M., and KOHN, W., 1958, *Phys. Rev.*, **109**, 892 (7).
MCCONNELL, H. M., 1958, *J. Chem. Phys.*, **28**, 1188 (5).
MACKE, W., 1955 a, *Phys. Rev.*, **100**, 992 (5); 1955 b, *Ann. Phys., Lpz.*, **17**, 1 (5).
MCLENNAN, J. A., 1959, *Phys. Rev.*, **115**, 1405 (7).
MCWEENY, R., 1955, *Proc. Roy. Soc.* A, **232**, 114 (5); 1956 a, *Proc. Roy. Soc.* A, **235**, 355 (5); 1956 b, *Proc. Roy. Soc.* A, **237**, 496 (5); 1957, *Proc. Roy. Soc.* A, **241**, 239 (5); 1959 a, *Proc. Roy. Soc.* A, **253**, 242 (5); 1959 b, *Phys. Rev.*, **114**, 1528 (5); 1960, *Rev. Mod. Phys.*, **32**, 335 (1, 5).
MCWEENY, R., and MIZUNO, Y., 1961, *Proc. Roy. Soc.* A, **259**, 554 (5).
MCWEENY, R., and OHNO, K. A., 1960, *Proc. Roy. Soc.* A, **255**, 367 (5).
MARCH, N. H., 1957, *Advanc. Phys.*, **6**, 1 (5).
MARCH, N. H., and YOUNG, W. H., 1958, *Proc. Phys. Soc.*, **72**, 182 (5).
MARGENAU, H., 1950, *Nature of Physical Reality* (New York: McGraw-Hill) (10).
MARTIN, P. C., and SCHWINGER, J., 1959, *Phys. Rev.*, **115**, 1342 (1, 3, 5).
MATSUBARA, T., 1955 a, *Prog. Theor. Phys., Japan*, **13**, 628 (6); 1955 b, *Prog. Theor. Phys., Japan*, **14**, 351 (1, 3, 4, 5, 6).
MATTIS, D. C., and BARDEEN, J., 1958, *Phys. Rev.*, **111**, 412 (7).
MAYER, J. E., 1955, *Phys. Rev.*, **100**, 1579 (5, 6).
MAZUR, P., and OPPENHEIM, I., 1957, *Physica*, **23**, 216 (4).
MIGDAL, A. B., 1957, *J. Exp. Theor. Phys., U.S.S.R.*, **32**, 399 (*Soviet Phys., JETP*, **5**, 333) (3).
MIZUNO, Y., and ISUYAMA, T., 1957, *Prog. Theor. Phys., Japan*, **18**, 33 (5, 6).
MONTROLL, E. W., and WARD, J. C., 1958, *Phys. Fluids*, **1**, 55 (3, 4); 1959, *Physica*, **25**, 423 (1, 3, 4, 7).
MORI, H., 1956, *J. Phys. Soc., Japan*, **11**, 1029 (7, 9); 1958, *Phys. Rev.*, **112**, 1829 (7); 1959, *Phys. Rev.*, **115**, 298 (7).
MORI, H., and ONO, S., 1952, *Progr. Theor. Phys., Japan*, **8**, 327 (7).
MORI, H., and ROSS, J., 1958, *Phys. Rev.*, **109**, 1877 (7).
MORSE, P. M., and FESHBACH, H., 1953, *Methods of Theoretical Physics* (London: McGraw-Hill) (3).
NAKAJIMA, S., 1955, *Advanc. Phys.*, **4**, 363 (4, 6); 1956, *Proc. Phys. Soc.* A, **69**, 441 (6, 7); 1959, see Varenna 1959, p. 275 (7).
NAKANO, H., 1956, *Progr. Theor. Phys., Japan*, **15**, 77 (1, 7); 1957, *Progr. Theor. Phys., Japan*, **17**, 145 (7); 1959, *Progr. Theor. Phys., Japan*, **22**, 453 (7); 1960 a, b, c, *Progr. Theor. Phys., Japan*, **23**, 180, 182, 526 (7); 1960 d, *Progr. Theor. Phys., Japan*, **23**, 527 (9).
NAQVI, M. A., 1959, *Nucl. Phys.*, **10**, 256 (1, 5).
VON NEUMANN, J., 1927 a, b, c, *Nachr. Ges. Wiss. Göttingen*, 1, 245, 273 (1); 1932, *Mathematische Grundlagen der Quantenmechanik* (Berlin: Springer) (Mathematical Foundations of Quantum Mechanics, Princeton University Press, 1955) (1, 10).
NISHIKAWA, K., 1960, *J. Phys. Soc., Japan*, **15**, 78 (7).
NITSOVICH, M. V., 1959, *Fiz. Met. Metallov.*, **7**, 641 (*Phys. Met. Metallogr.*, **7**, No. 5, 1) (6).

ONO, S., 1953, *Proc. Kyoto and Tokyo Conf. Theor. Phys.*, 449 (7); 1954, *Prog. Theor. Phys., Japan*, **12**, 113 (7); 1958, see Prigogine 1958, p. 229 (6, 7).
OPECHOWSKI, W., 1953, *Rev. Mod. Phys.*, **25**, 264 (9).
OPPENHEIM, I., and MAZUR, P., 1957, *Physica*, **23**, 197 (4).
OPPENHEIM, I., and ROSS, J., 1957, *Phys. Rev.*, **107**, 28 (4).
OSAKA, Y., 1959, *Prog. Theor. Phys., Japan*, **22**, 437 (6).
PARRENT, G. B., and ROMAN, P., 1960, *Nuovo Cim.*, **15**, 370 (8).
PAVLIKOVSKIĬ, A., and SHCHURUVNA, V., 1958, *Dokl. Akad. Nauk, S.S.S.R.*, **118**, 61 (*Soviet Phys., Doklady*, **3**, 71) (4); 1959, *Dokl. Akad. Nauk, S.S.S.R.*, **124**, 69 (*Soviet Phys., Doklady*, **4**, 95) (4).
PEACOCK, T. E., and MCWEENY, R., 1959, *Proc. Phys. Soc.*, **74**, 385 (5).
PENROSE, O., 1960, *Proc. Roy. Soc.* A, **256**, 106 (5).
PITAEVSKIĬ, L. P., 1959, *J. Exp. Theor. Phys., U.S.S.R.*, **36**, 1168 (*Soviet Phys., JETP*, **9**, 830) (3).
PRANGE, R., and KLEIN, A., 1958, *Phys. Rev.*, **112**, 1008 (3).
PRIGOGINE, I. (editor), 1958, *Transport Processes in Statistical Mechanics* (New York: Interscience) (1, 7).
PRIGOGINE, I., and BALESCU, R., 1959 a, b, *Physica*, **25**, 281, 302 (7).
PRIGOGINE, I., and ONO, S., 1959, *Physica*, **25**, 171 (7).
PRYCE, M. H. L., and WARD, J. C., 1947, *Nature, Lond.*, **160**, 435 (8).
REDFIELD, A. G., 1957, *I.B.M. J. Res. Developm.*, **1**, 19 (9); 1959, *Phys. Rev.*, **116**, 315 (9).
RIESENFELD, W. B., and WATSON, K. M., 1956, *Phys. Rev.*, **104**, 492 (4).
ROMAN, P., 1959, *Nuovo Cim.*, **13**, 974 (8).
VON ROOS, O., 1960, *J. Math. Phys.*, **1**, 107 (7).
ROSS, J., and KIRKWOOD, J. G., 1954, *J. Chem. Phys.*, **22**, 1094 (7).
SAITÔ, N., 1960, *Phys. Rev.*, **117**, 1163 (7).
SAKAI, 1937, *Proc. Math. Phys. Soc., Japan*, **19**, 172 (10).
SALAM, A., 1953, *Prog. Theor. Phys., Japan*, **9**, 550 (3, 6).
SALPETER, E. E., 1958, *Ann. Phys., New York*, **5**, 183 (4).
SAROLÉA, L., and KOPPE, H., 1958, *Z. Phys.*, **151**, 385 (5).
SAWICKI, J., 1960, *Prog. Theor. Phys., Japan*, **24**, 213 (3).
SCHAFROTH, M. R., 1951, *Helv. Phys. Acta*, **24**, 645 (6).
SCHAFROTH, M. R., BUTLER, S. T., and BLATT, J. M., 1957, *Helv. Phys. Acta*, **30**, 93 (4, 6).
SCHRÖDINGER, E., 1948, *Statistical Thermodynamics* (Cambridge: University Press) (1).
SCHUMACHER, R. T., 1958, *Phys. Rev.*, **112**, 837 (9).
SEIDEN, J., 1956 a, b, *C.R. Acad. Sci., Paris*, **242**, 454, 763 (9).
SHER, A., and PRIMAKOFF, H., 1960, *Phys. Rev.*, **119**, 178 (9).
SHULL, H., 1959, *J. Chem. Phys.*, **30**, 1405 (5).
SIEGERT, A. J. F., 1960, *Physica*, **26**, S30 (4).
SIEGERT, A. J. F., and TERAMOTO, E., 1958, *Phys. Rev.*, **110**, 1232 (4).
SKROTSKIĬ, G. V., and KOKIN, A. A., 1959, *J. Exp. Theor. Phys., U.S.S.R.*, **36**, 169 (*Soviet Phys., JETP*, **9**, 116) (9).
SNYDER, H. S., PASTERNACK, S., and HORNBOSTEL, J., 1948, *Phys. Rev.*, **73**, 440 (8).
SOLOMON, I., 1958, *Phys. Rev.*, **110**, 61 (9).
SONDHEIMER, E. H., and WILSON, A. H., 1951, *Proc. Roy. Soc.* A, **210**, 173 (6).
STILLINGER, F. H., and KIRKWOOD, J. G., 1960, *Phys. Rev.*, **118**, 361 (4).
STOKES, G. G., 1852, *Trans. Camb. Phil. Soc.*, **9**, 399 (8).
SUHL, H., 1958, *J. Phys. Chem. Solids*, **4**, 278 (9).
TAKABAYASI, T., 1954, *Prog. Theor. Phys., Japan*, **11**, 341 (4).
THOULESS, D. J., 1957, *Phys. Rev.*, **107**, 1162 (4); 1959, *Phys. Rev.*, **116**, 21 (4); 1961, *Nucl. Phys.*, in the press (5).
TODA, M., 1958, *J. Phys. Soc. Japan*, **13**, 1266 (7).
TOLHOEK, H. A., 1956, *Rev. Mod. Phys.*, **28**, 277 (8).
TOLHOEK, H. A., and COX, J. A. M., 1953, *Physica*, **19**, 101 (8).
TOLHOEK, H. A., and DE GROOT, S. R., 1951 a, b, c, *Physica*, **17**, 1, 17, 81 (8).
TOLMAN, R. C., 1938, *The Principles of Statistical Mechanics* (Oxford: University Press) (1, 2, 4, 10).

TOMITA, K., 1958 a, *Prog. Theor. Phys., Japan*, **19**, 541 (9); 1958 b, *Prog. Theor. Phys., Japan*, **20**, 743 (9).
TREDGOLD, R. H., 1957, *Phys. Rev.*, **105**, 1421 (5, 6).
VALATIN, J. G., 1958, *Nuovo Cim.*, **7**, 843 (6).
VAN HOVE, L., 1955, *Physica*, **21**, 517 (3, 7); 1957, *Physica*, **23**, 441 (3, 10); 1959, *Physica*, **25**, 268 (10).
VARENNA, 1959, *Proceedings of the 'Enrico Fermi' International Summer School on Thermodynamics of Irreversible Processes* (Bologna: Zanichelli) (7).
VEDENOV, A. A., 1959, *J. Exp. Theor. Phys., U.S.S.R.*, **36**, 641 (*Soviet Phys., JETP*, **9**, 446) (3).
VEDENOV, A. A., and LARKIN, A. I., 1959, *J. Exp. Theor. Phys., U.S.S.R.*, **36**, 1133 (*Soviet Phys., JETP*, **9**, 806) (3).
WAKITA, H., 1960, *Prog. Theor. Phys., Japan*, **23**, 32 (10).
WANGSNESS, R. K., 1956 a, *Phys. Rev.*, **101**, 1 (9); 1956 b, *Phys. Rev.*, **104**, 857 (9).
WANGSNESS, R. K., and BLOCH, F., 1953, *Phys. Rev.*, **89**, 728 (9).
WANNIER, G. H., 1959, *Elements of Solid State Theory* (Cambridge: University Press), Ch. 7 (9).
WATSON, K. M., 1956, *Phys. Rev.*, **103**, 489 (3, 4).
VON WEIZSÄCKER, C. F., 1935, *Z. Phys.*, **96**, 431 (5).
WICK, G. C., 1950, *Phys. Rev.*, **80**, 268 (4).
WIGNER, E., 1932, *Phys. Rev.*, **40**, 749 (7).
WILCOX, L. R., and LAMB, W. E., 1960, *Phys. Rev.*, **119**, 1915 (9).
WILSON, A. H., 1953, *Theory of Metals* (Cambridge: University Press) (1).
WOLF, E., 1954, *Nuovo Cim.*, **12**, 884 (8); 1959 a, *Nuovo Cim.*, **13**, 1165 (8); 1959 b, *Proc. Phys. Soc.*, **74**, 269 (8); 1960, *Proc. Phys. Soc.*, **76**, 424 (8).
YAGLOM, A. YA., 1956, *Teor. Veroyatn. Primenen.*, **1**, 161 (4).
YAMAMOTO, T., 1960, *Phys. Rev.*, **119**, 701 (6).
YATSIV, S., 1959, *Phys. Rev.*, **113**, 1522 (9).
YOKOTA, T., 1960, *J. Phys. Soc., Japan*, **15**, 779 (7).
YOUNG, W. H., and MARCH, N. H., 1960, *Proc. Roy. Soc.* A, **256**, 62 (5, 6).
YVON, J., 1957, *Nucl. Phys.*, **4**, 1 (4); 1958, *C.R. Acad. Sci., Paris*, **246**, 2850; 1960 a, *J. Phys. Radium*, **21**, 505 (4, 9); 1960 b, *J. Phys. Radium*, **21**, 569 (4).
ZAIDI, M. H., 1959, *Phys. Rev.*, **116**, 241 (8).
ŽELAZNY, R., 1960, *Phys. Rev.*, **117**, 1 (7).
ZIGENLAUB, R., 1959, *Fiz. Tverd. Tel.*, **1**, 1053 (*Soviet Phys., Solid State*, **1**, 964) (7).
ZUBAREV, D. N., 1954, *Dokl. Akad. Nauk, S.S.S.R.*, **95**, 757 (4, 6); 1960, *Usp. Fiz. Nauk*, **71**, 71 (*Soviet Phys., Uspekhi*, **3**, 320) (1, 3, 7).

THE DYNAMICS OF HIGH TEMPERATURE PLASMAS

By W. B. THOMPSON

Theoretical Physics Division, Atomic Energy Research Establishment, Harwell, Berks.

CONTENTS

	PAGE
Abstract	363
Introduction—The plasma and plasma models	364
§ 1. The magnetohydrodynamic representation of the plasma	365
1.1 The magnetohydrodynamic equations	366
1.2. Conditions for idealized magnetohydrodynamics	367
1.3. Magnetohydrodynamic equilibrium	368
1.4. Small oscillations	370
1.5. Stability	372
§ 2. The kinetic theory of a dense plasma	376
§ 3. The kinetic theory of the diffuse unmagnetized plasma	384
3.1. The Vlasov–Landau theory of high frequency phenomena	384
3.2. Dielectric behaviour and the screened field	388
3.3. The Fokker–Planck equation and dielectric screening	390
3.4. Electrostatic instabilities	395
§ 4. The kinetic theory of the diffuse plasma in a magnetic field	397
4.1. The dielectric properties of a cold magnetized plasma	397
4.2. Dielectric properties of a hot magnetized plasma	400
4.3. Plasma in a non-uniform field ; approximate methods	403
4.4. Some exact solutions for the steady state	406
4.5. Approximate procedure for solving the Vlasov equation	408
4.6. The magnetohydrodynamic equations	410
4.7. The equations of state	411
4.8. Transverse transport coefficients	411
4.9. Solutions to the zero-order constraint ; the guiding centre distribution	412
4.10. Perturbation theory	414
4.11. Theory of sound	418
§ 5. Conclusions	420
Acknowledgments	422
References	422

Abstract. A survey is given of present models of the dynamic behaviour of a fully ionized plasma, with particular emphasis on those useful in stability studies. The models discussed include magnetohydrodynamics which, while not fully justified, is widely used, the classical kinetic theory which probably is not applicable to hot laboratory plasmas, and the collisionless kinetic theory, which while incomplete seems suited to the discussion of the stability of plasmas. The properties of small oscillations are discussed, both sound waves and electrical oscillations, and stress is laid on the dielectric behaviour of the plasma. A consistent derivation of the Fokker–Planck equation is sketched, using the dielectric properties of the plasma. Many important dynamical problems are omitted—there is no discussion of the collisionless shock, or of the breakdown of hydrodynamic behaviour through electron runaway in electric fields, and there is little discussion of particular configurations.

Introduction—The Plasma and Plasma Models

Although the study of ionized gases has a long history, the actual dynamics of the ionized gas, or plasma, has been strangely neglected until recently. This has been largely a result of the very fruitfulness of experiments with ionized gases, which became a most important ancestor of modern atomic and nuclear physics. The behaviour of most laboratory plasmas is dominated by inelastic collisions between atoms, leading to ionization, dissociation or excitation, and experimenters rightly neglected the study of plasma dynamics in favour of investigations of atomic structure. A further reason for the neglect of plasma dynamics is that these inelastic collision processes, while fascinating in themselves, render the dynamics of the partially ionized gas quantitatively unintelligible—except in the simplest circumstances. On the other hand, the completely ionized plasma, a fairly simple object for theoretical study, has been until recently almost unrealizable in the laboratory.

During the past decade there has been a spectacular revival of interest in this neglected physical problem, the stimulus coming from widely different sources. One of these has been radio physics, which has required an understanding of at least the electrical properties of the ionized layers in the atmosphere (Ratcliffe 1959). Since Hale's measurement of the magnetic fields of sun spots (1908), solar astronomers (see Alfvén 1950) have had an interest in the subject, while the discovery of highly magnetized variable stars by Babcock (1947) and the polarization of starlight by Hiltner (1950) has extended the range of plasma physics throughout the galaxy. If the more recent identification of much cosmic radio noise with the cyclotron radiation of electrons moving in magnetic fields (Shklovsky 1953) is permanent, plasma physics will be even more closely interwoven with astrophysics.

Perhaps an even stronger stimulus towards research in plasma dynamics has been provided by its technological promises. Of these, the most spectacular is the hope of the controlled release of energy from thermonuclear reactions (Post 1956, Thompson 1957). There is no doubt that a mass of deuterium and tritium raised to a sufficiently high temperature could release not only large amounts of energy but enough neutrons to make tritium breeding feasible. If such high temperatures can be reached, a further revolution in primary power supplies is possible; however, since the temperatures are extreme ($\sim 8 \times 10^7 °\text{K} \sim 8$ kev), some non-material method of confining and thermally isolating the gas is needed and at present containment of the hot, ionized gas in properly shaped magnetic fields is the most promising approach (Post 1956, Thompson 1957). Considerable efforts have been made to produce a magnetically confined plasma with, as yet, a disappointing lack of success (Artsimovich 1958, 1959, Tuck 1958, 1959, Thonemann 1958, 1959). Most of the obvious containing geometries are destroyed by dynamic instabilities which not only allow the plasma to escape from confinement but destroy the symmetry and intelligibility of the system (Thonemann 1958, 1959). The study of plasma stability, forced upon us by the unco-operative behaviour of magnetic bottles, has greatly increased our knowledge of the dynamics of plasma in magnetic fields although not yet to the state where the thermonuclear problem is immediately solvable.

In addition to their possible practical importance, plasma dynamic studies are of intrinsic interest to the physicist on several levels. Macroscopically, the coupling between the motions of a conducting gas and variations in the electromagnetic field produces the subject of magnetohydrodynamics, a vast unexplored extension of hydrodynamics. Microscopically, the plasma consists of an assembly of particles interacting through weak, long range forces ; thus it poses a many-body problem of great interest. Finally, the bridge between the microscopic and macroscopic descriptions of a plasma, particularly in a magnetic field, has certain unusual and interesting features.

One rather distressing aspect of this subject is the lack of a large body of experimentally confirmed theory ; indeed the connection between theory and experiment is sufficiently tenuous for the entire body of the theory to be suspect. The theory so far has been developed only in a rather idealized form, which, it is felt, should describe a sufficiently hot plasma, to which inelastic processes are irrelevant; on the other hand, experiments are conducted with real plasmas, cooled by walls and distorted into asymmetric motion by instabilities. In this, the subject may resemble hydrodynamics where at high Reynolds number, when viscous stresses become apparently negligible, the flow becomes unstable and turbulent so that idealized hydrodynamics has a limited application.

In spite of these drawbacks, this article will be concerned with the idealized plasma, and will be devoted to a brief exposition of the theoretical models that have been studied in an attempt to understand plasma dynamics. In the first section, the magnetohydrodynamic representation will be described; in the second, the basis of this representation in classical kinetic theory will be examined and its limitations noted. The third section introduces an alternative kinetic description which is used to discuss the electromagnetic behaviour of a plasma, and to provide a basis for the classical kinetic theory. Finally in the last section the dynamic consequences of this theory will be described. Throughout, the bias will be towards the study of plasma stability, partly because of the linear character of this problem, and partly because of my own interest in thermonuclear research.

§ 1. The Magnetohydrodynamic Representation of the Plasma

Before turning to the simplest representation of plasma dynamics, it is worth recalling the gross electrical properties of a plasma. These can be summarized by the frequency dependent dielectric coefficient $\epsilon(\omega) = 1 - \omega_0^2/\omega^2$, where $\omega_0 = (4\pi ne^2/m)^{1/2}$ is the Langmuir plasma frequency, typically about 10^3 Mc/s. For frequencies much below ω_0, the plasma behaves as a good electrical conductor, with a resistivity η falling rapidly with increasing temperature ($\eta \sim T^{-3/2}$). Moreover, the plasma has an overwhelming tendency to remain electrically uncharged, any excess electrical charge disappearing in a time of order ω_0^{-1}. In dealing with low frequency problems this condition of quasi-neutrality is frequently invoked, and it is important that its meaning be clear. We shall insist that, if Z is the ionic charge, $Zn_+ = n_-$ is a good approximation for the number densities, but at the same time we will allow $\operatorname{div} \mathbf{E} \neq 0$; thus although in a plasma the number of ionic and electronic charges are almost equal, a small residual charge separation

is enough to produce a significant electric field. This approximation implies that the electric field in a plasma usually has a rather subsidiary role and is determined by the flow.

1.1. *The Magnetohydrodynamic Equations*

In view of these electrical properties, a useful model of the plasma in interaction with a magnetic field consists of an electrically conducting ideal gas interacting with a magnetic field (Cowling 1957). To complete this model it is necessary to use some constitutive equation relating the current to the fields and the motion. Usually it is enough to consider a simple local Ohm's law, but using the local electric field, i.e. the force acting on a unit charge moving with the velocity $\mathbf{v}$ of the fluid ; thus,

$$\eta \mathbf{j} = \mathbf{E} + \mathbf{v} \times \mathbf{B}, \qquad \ldots\ldots(1.1)$$

where η is the resistivity. Often it is sufficient to consider the conductivity as perfect, i.e. $\eta = 0$.

The equation of motion is the usual hydrodynamic equation with the ponderomotive force $\mathbf{j} \times \mathbf{B}$ as a driving force,

$$\rho \frac{D}{Dt}\mathbf{v} = -\nabla p + \mathbf{j} \times \mathbf{B}, \qquad \ldots\ldots(1.2)$$

where

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v}.\nabla, \quad \text{as usual.}$$

To these must be added Maxwell's equations, or, since the displacement current is usually omitted, the pre-Maxwell equations,

$$\operatorname{curl} \mathbf{B} = 4\pi \mathbf{j}, \quad \operatorname{div} \mathbf{B} = 0, \quad \operatorname{curl} \mathbf{E} = -\partial \mathbf{B}/\partial t. \qquad \ldots\ldots(1.3)$$

The current and the electric field may be eliminated from these, to leave

$$\rho \frac{D}{Dt}\mathbf{v} = -\nabla\left(p + \frac{B^2}{8\pi}\right) + \left(\frac{\mathbf{B}}{4\pi}.\nabla\right)\mathbf{B}, \qquad \ldots\ldots(1.4)$$

$$\frac{D}{Dt}\mathbf{B} + \mathbf{B} \operatorname{div} \mathbf{v} - (\mathbf{B}.\nabla)\mathbf{v} = 4\pi\eta\nabla^2 \mathbf{B}. \qquad \ldots\ldots(1.5)$$

To derive (1.4) substitute $\mathbf{j}$ from (1.3) into (1.2) and expand $\operatorname{curl} \mathbf{B} \times \mathbf{B}$. For (1.5) take the curl of (1.2), substitute for $\operatorname{curl} \mathbf{j}$ and $\operatorname{curl} \mathbf{E}$ from (1.3) and expand $\operatorname{curl} \operatorname{curl} \mathbf{B}$ and $\operatorname{curl}(\mathbf{v} \times \mathbf{B})$.

To complete the set, we need the equation of continuity

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \mathbf{v}) = 0, \qquad \ldots\ldots(1.6)$$

and some equation of state. A popular choice is that for adiabatic flow,

$$\frac{D}{Dt}(p\rho^{-\gamma}) = 0, \qquad \ldots\ldots(1.7)$$

although the fluid is often considered incompressible, $\operatorname{div} \mathbf{v} = 0$, and occasionally attempts have been made to use a full equation of state together with an energy transport equation. Logically this last procedure is essential if there is a finite resistivity.

1.2. *Conditions for Idealized Magnetohydrodynamics*

For most dynamical problems even this highly idealized and relatively simple approach is too complex, and it is simplified by setting $\eta = 0$, i.e. considering the plasma as a perfect conductor, in which limit the Eqns (1.1) to (1.7) are consistent. The applicability of this idealization to any particular situation may be assessed by comparing the magnitude of the convective terms on the left of (1.5) with the dissipative term on the right. A rough assessment is obtained by writing $\mathbf{v}.\nabla\mathbf{B} \simeq vB/L$, and $4\pi\eta\nabla^2\mathbf{B} \simeq 4\pi\eta(B/L^2)$, and forming the dimensionless magnetic Reynolds number

$$R_M = 4\pi vL/\eta. \qquad \ldots\ldots(1.8)$$

If R_M is large, convection dominates over conduction and as R_M increases, the ideally conducting model becomes increasingly valid.

If, in addition, the flow is to be determined by the magnetic fields, then from (1.4)

$$\frac{\rho v^2}{L} \simeq \frac{1}{L}\frac{B^2}{4\pi}; \quad \text{or} \quad v \simeq \sqrt{\left(\frac{B^2}{4\pi\rho}\right)} = V_A,$$

the Alfvén speed, and the condition for ' typical ' magnetohydrodynamic behaviour becomes

$$R_M = B\sqrt{\left(\frac{4\pi}{\rho}\right)}\frac{L}{\eta}. \qquad \ldots\ldots(1.9)$$

Typical values are displayed in Table 1.

Table 1. The Magnetic Reynolds Number R_M

System	Scale	Field (gauss)	Density (g cm^{-3})	η^{-1} (sec^{-1})	R_M
Astrophysical					
Earth's core	10^3 km	1	12	10^{15}	10^3
Ionosphere	10^3 km	0·2	10^{-8}	10^{11}	10
Solar atmosphere	10^4 km	10^2	10^{-9}	10^{13}	10^8
Solar corona	10^6 km	10^{-5}	10^{-20}	10^{18}	10^{11}
Hot gaseous nebula	0·1 light year	10^{-5}	10^{-22}	10^{13}	10^{14}
Hot interstellar gas	1·0 light year	10^{-6}	10^{-24}	10^{13}	10^{15}
Laboratory					
Mercury	10 cm	1000	13·5	10^{16}	0·1
Liquid sodium	10 cm	1000	0·93	10^{17}	4·0
Normal discharge plasma	10 cm	1000	10^{-6}	10^{15}	10^3
Hot confined plasma ($n = 10^{15}$ cm^{-3} $T = 10^6$ °K)	10 cm	1000	10^{-9}	10^{13}	10^4

The great simplification introduced by idealized magnetohydrodynamics arises through the appearance of several useful conservation theorems. For example, from (1.7) conservation of entropy holds, while (1.5) which, with the right-hand side equal to 0, has the form of the equation of transport of vorticity, implies that the flux threading any closed loop lying within the fluid is conserved, an immediate consequence of Faraday's induction law. Conservation of energy is a little more

delicate. Since the forces are in general not derivable from a potential the Bernoulli integral is not conserved. On the other hand, it is readily shown that there exists a conserved integral over the volume of the fluid

$$\epsilon = \int \left(\tfrac{1}{2}\rho v^2 + \frac{B^2}{8\pi} + \frac{p}{\gamma - 1} \right) d\tau. \qquad \ldots\ldots (1.10)$$

If the surface of the fluid is not crossed by field lines, which is a necessary condition for magnetohydrodynamic equilibrium, the integral ϵ over the volume of the fluid is constant, while otherwise its rate of change is $\int ds (B^2/8\pi) v_n$, where $v_n = \mathbf{v} \cdot \mathbf{B}/|B|$, which represents the rate at which magnetic energy is swept in from the vacuum.

1.3. *Magnetohydrodynamic Equilibrium*

In view particularly of its possible application in thermonuclear research, there is great interest in the existence of configurations of magnetohydrostatic equilibrium. If the pressure is a scalar, there is a stringent condition, first observed by Ferraro (1937), on those magnetic configurations which can support a pressure, for these must satisfy

$$\begin{aligned} \mathbf{j} \times \mathbf{B} &= \nabla p, \\ \operatorname{curl} (\mathbf{j} \times \mathbf{B}) &= 0. \end{aligned} \qquad \ldots\ldots (1.11)$$

From (1.11) it is clear that $\mathbf{B} \cdot \nabla p = 0$ and $\mathbf{j} \cdot \nabla p = 0$, hence the field and current are everywhere tangential to surfaces of constant pressure and the field and current lines must cover these surfaces, the magnetic surfaces. In order to confine a gas, the magnetic surfaces must be nested, and in the simplest confining system form a set of nested toroids (Kruskal and Kulsrud 1958), a topology which is exploited in a large number of experimental or speculative thermonuclear devices: the toroidal pinched discharge, the stellarator, Astron.

Several classes of analytic representations of confining equilibria have been given, usually for cylindrically symmetric systems (Kruskal and Kulsrud 1958, Laing *et al.* 1959, Lüst and Schlüter 1957). In these, the r, z components of the field described in a cylindrical coordinate system may be derived from a stream function ψ

$$B_r = \frac{1}{r} \frac{\partial \psi}{\partial z}(r, z), \quad B_z = -\frac{1}{r} \frac{\partial \psi}{\partial r},$$

which automatically satisfies the condition $\operatorname{div} \mathbf{B} = 0$. The second of the equilibrium equations

$$\begin{aligned} \frac{\partial p}{\partial \phi} = 0 &= B_r \frac{\partial}{\partial r}(rB_\phi) + B_z \frac{\partial}{\partial z}(rB_\phi) = \frac{1}{r}\left[\frac{\partial \psi}{\partial z} \frac{\partial}{\partial r}(rB_\phi) - \frac{\partial \psi}{\partial r} \frac{\partial}{\partial z}(rB_\phi) \right] \\ &= \frac{1}{r} J(\psi, rB_\phi;\ r, z) = 0 \end{aligned} \qquad \ldots\ldots (1.12)$$

indicates that rB_ϕ is a function of ψ alone, $rB_\phi = f(\psi)$, while the condition $\mathbf{B} \cdot \nabla p = 0$, or

$$\frac{1}{r} \frac{\partial \psi}{\partial z} \frac{\partial p}{\partial r} - \frac{1}{r} \frac{\partial \psi}{\partial r} \frac{\partial p}{\partial z} = 0 \qquad \ldots\ldots (1.13)$$

implies p is also a function of ψ; $p = g(\psi)$.

The remaining equilibrium equations take the form

$$r\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial \psi}{\partial r}\right)+\frac{\partial^2\psi}{\partial z^2}+ff'+r^2g'=0, \qquad \ldots\ldots(1.14)$$

where the prime denotes differentiation with respect to ψ.

The last equation is completely general, but at this stage a choice must be made for the arbitrary functions f and g. If the former is taken as linear and the latter as either linear or quadratic in ψ, the equation (1.14) becomes linear, and may be solved, although the application of realistic boundary conditions is difficult. In particular, if the pressure is negligible the fields are force-free, and the configuration is represented by

$$\psi = A\cos kz\,.\,rB_1[(b^2-k^2)^{1/2}r] \qquad \ldots\ldots(1.15)$$

where $f=b\psi$, A and k are arbitrary and B_1 is a Bessel function.

Linear combinations of the solution (1.15) may be used to represent toroidal confining fields and since a perfectly conducting wall can be placed on any surface ψ = constant, quite reasonable representations of fields are easily obtained.

An alternative approach to the equilibrium of toroidal configurations is to consider an equilibrium field in an infinite cylinder, and to ask how this is deformed as the cylinder is bent into a torus (Laing *et al.* 1959). This has been done by an expansion in inverse powers of the aspect ratio of the torus. In zero order,

$$\psi_0 = J_0(\kappa\bar{r}), \qquad \ldots\ldots(1.16)$$

and

$$\psi_1 = \tfrac{1}{2}\cos\theta[J_0(\kappa\bar{r})-\frac{\alpha}{\kappa}\bar{r}^2J_1(\kappa\bar{r})+\bar{A}J_1(\kappa\bar{r})]$$

where $\alpha\psi^2$ is the pressure and $\bar{r}$ and θ are polar coordinates in the circular cross section of the torus.

Measurements made of the magnetic fields in ZETA (Lees and Rusbridge 1960) have shown that in the current bearing region these are surprisingly well fitted by $J_0(\kappa r)$, where distances are measured from the magnetic axis, i.e. the limit line of the nested toroidal surfaces.

A second interesting class is that of magnetic equilibria bounded by spherical surfaces. Of these the simplest is the magnetic analogue of Hill's spherical vortex (Whipple 1960) for which, in spherical polar coordinates

$$\begin{aligned}
\psi &= \frac{B_0}{2}\sin^2\theta\left(R^2-\frac{A^3}{R}\right), \quad R>A,\\
&= -\tfrac{3}{4}B_0\sin^2\theta R^2\left(1-\frac{R^2}{A^2}\right), \quad R<A,\\
p &= \frac{45}{8}\frac{B_0^2}{A^2}\sin^2\theta R^2\left(1-\frac{R^2}{A^2}\right), \quad R<A, \qquad \ldots\ldots(1.17)\\
&= 0, \qquad\qquad R>A.
\end{aligned}$$

It is possible to include B_ϕ in those spherical objects, and (1.12) implies in spherical coordinates

$$B_\phi = \frac{1}{R\sin\theta}f(\psi).$$

If $f(\psi) = \mu\psi, \psi < 0$; $g(\psi) = -\lambda\psi, \psi < 0$ and f and $g = 0$ for $\psi > 0$; then the equilibrium equations are satisfied by $\psi = F(R)\sin^2\theta$

$$F = \frac{\lambda}{\mu^2}R^2 - \nu\left(\frac{\sin\mu R}{\mu R} - \cos\mu R\right) \qquad \ldots\ldots(1.18)$$

where

$$\lambda = \tfrac{3}{2}B_0\mu^2\frac{\sin\mu A - \mu A\cos\mu A}{(3-\mu^2A^2)\sin\mu A - 3\mu A\cos\mu A} \qquad \ldots\ldots(1.19)$$

and

$$\nu\left(\frac{\sin\mu A}{\mu A} - \cos\mu A\right) = \frac{\lambda}{\mu^2}A^2.$$

At $\mu A = 5{\cdot}763$, λ and ν become infinite although λ/ν remains finite. Thus if the external field is derived from $\psi = B_0\sin^2\theta(R^2 - A^3/R)$ the field and pressure within the sphere remain finite as the external field is decreased to zero, whereupon the plasma appears in self-confined equilibrium, unfortunately with a negative pressure.

Within the context of ideal magnetohydrodynamics, surface currents and field discontinuities are permitted and there is a further class of configurations exploiting these, the magnetohydrodynamic analogues of the magnetic cusps and the magnetic mirror.

1.4. *Small Oscillations*

The simplest dynamical problem of much interest is the propagation of small oscillations through a magnetohydrodynamic system. To study these, the equations (1.4) to (1.7) may be linearized about a static equilibrium. Further simplification is possible when it is observed that not only the velocities $\mathbf{v}(x, t)$ are small, but also the displacements $\boldsymbol{\xi}(x, t)$, so that $\dot{\boldsymbol{\xi}} \simeq \partial\boldsymbol{\xi}/\partial t = \mathbf{v}$. It is then possible to express the perturbed density ρ', the perturbed pressure p', and the perturbed magnetic field $\mathbf{B}'$ in terms of the displacements $\boldsymbol{\xi}$ and the unperturbed quantities ; (1.6) yields on time integration

$$\rho' + \operatorname{div}(\rho_0\boldsymbol{\xi}) = 0\ ;\quad \rho' = -\operatorname{div}(\rho_0\boldsymbol{\xi}) = -(\boldsymbol{\xi}.\nabla\rho_0 + \rho_0\operatorname{div}\boldsymbol{\xi}) \qquad \ldots\ldots(1.20)$$

and (1.7) yields

$$\left(\frac{Dp}{Dt}\right)' - \gamma\frac{p_0}{\rho_0}\left(\frac{D\rho}{Dt}\right)' = 0,$$

i.e.

$$p' = -\boldsymbol{\xi}.\nabla p_0 - \gamma p_0\operatorname{div}\boldsymbol{\xi} \qquad \ldots\ldots(1.21)$$

while (1.5) with $\eta = 0$ yields

$$\mathbf{B}' = (\mathbf{B}_0.\nabla)\boldsymbol{\xi} - (\boldsymbol{\xi}.\nabla)\mathbf{B}_0 - \mathbf{B}_0\operatorname{div}\boldsymbol{\xi} \qquad \ldots\ldots(1.22)$$

and the equation of motion becomes

$$\rho_0\ddot{\boldsymbol{\xi}} = -\nabla p' + (\operatorname{curl}\mathbf{B}')\times\mathbf{B}_0 + (\operatorname{curl}\mathbf{B}_0)\times\mathbf{B}' \qquad \ldots\ldots(1.23)$$

$$= \nabla(\boldsymbol{\xi}.\nabla p_0 + \gamma p_0\operatorname{div}\boldsymbol{\xi}) - \frac{1}{4\pi}\{\nabla(\mathbf{B}_0.\mathbf{B}') - (\mathbf{B}_0.\nabla)\mathbf{B}' - (\mathbf{B}'.\nabla)\mathbf{B}_0\}.$$

For the particular case of a uniform undisturbed system

$$\ddot{\boldsymbol{\xi}} = \frac{\gamma p_0}{\rho_0}\nabla(\nabla.\boldsymbol{\xi}) - \frac{1}{4\pi\rho_0}\nabla[(\mathbf{B}_0.\nabla)\mathbf{B}_0.\boldsymbol{\xi} - B_0^2 \operatorname{div}\boldsymbol{\xi}] + \frac{1}{4\pi\rho_0}\mathbf{B}_0.\nabla[(\mathbf{B}_0.\nabla)\boldsymbol{\xi} - \mathbf{B}_0 \operatorname{div}\boldsymbol{\xi}]. \quad \ldots\ldots(1.24)$$

By introducing a plane wave substitution $\boldsymbol{\xi} = \boldsymbol{\xi}_0 \exp[i2\pi(\mathbf{k}.\mathbf{x} - Vt)/\lambda]$, where $\mathbf{k}$ is a unit vector in the direction of propagation, λ is the wavelength and V the phase velocity, (1.16) is reduced to three linear homogeneous algebraic equations for the components of the wave amplitude $\boldsymbol{\xi}_0$. These equations have non-trivial solutions only if the determinant of the coefficients vanishes, and a cubic equation for the square of the phase velocity V is obtained. This is most easily obtained by considering not the equations for the components of $\boldsymbol{\xi}_0$ but the three having on the left $\mathbf{B}_0.\mathbf{k}\times\boldsymbol{\xi}$, $\mathbf{k}.\boldsymbol{\xi}$, $\mathbf{B}_0.\boldsymbol{\xi}$, whereupon the dispersion relation appears factorized into a linear and a quadratic term. The roots are

$$V_A = \pm C_A \cos\theta \quad \ldots\ldots(1.25)$$

$$V_+ = \pm\left\{\frac{C_A^2 + C_s^2 + [(C_A^2 + C_s^2)^2 - 4C_A^2 C_s^2 \cos^2\theta]^{1/2}}{2}\right\}^{1/2} \quad \ldots\ldots(1.26)$$

$$V_- = \pm\left\{\frac{C_A^2 + C_s^2 - [(C_A^2 + C_s^2) - 4C_A^2 C_s^2 \cos^2\theta]^{1/2}}{2}\right\}^{1/2} \quad \ldots\ldots(1.27)$$

where $C_A = B/\sqrt{(4\pi\rho)}$ is the Alfvén speed ; and $C_s = \sqrt{(\gamma p/\rho)}$ the sound speed and θ the angle $\cos^{-1}\mathbf{k}.\mathbf{B}/|\mathbf{B}|$ between the field and the normal $\mathbf{k}$ to the surfaces of constant phase. The first of these speeds is associated with the Alfvén wave, a purely transverse oscillation of the line of force. The second two waves involve compression of the fluid, and are the analogues of sound waves. While these are not in general longitudinal, i.e. $\boldsymbol{\xi} \nparallel \mathbf{k}$, they are coplanar with $\mathbf{B}$, i.e. $\boldsymbol{\xi}$, $\mathbf{k}$ and $\mathbf{B}$ lie in a plane and $\boldsymbol{\xi}.\mathbf{k}\times\mathbf{B} = 0$. For the fast wave V_+ the perturbations in magnetic pressure and in gas pressure act together, the compression being primarily transverse to the field, while for the slow wave V_- these pressures are opposed and compression is primarily along the field. V_+ ranges from $\sqrt{(C_A^2 + C_s^2)}$, the magneto-sonic speed across the field lines, to the greater of C_A, C_s, along the field lines ; while V_- ranges from the lesser of C_A, C_s along the field lines to zero across them.

The implications of a non-isotropic phase velocity are not immediately obvious. The curves showing $V(\theta)$ do not represent the wave front from a point source ; instead the radius vector represents the velocity of an infinite plane component, and to obtain the wave front from a point source it is necessary to construct the normals to the radius vector on $V(\theta)$ and find the envelope of those normals. When this is done, the Alfvén wave front reduces to two points moving with velocity $\pm C_A$ along the field lines, the fast wave front has a roughly elliptical shape, while the slow wave front is found to consist of two small cusped surfaces centred on points moving along the field lines with velocities $\pm \min(C_s, C_A)$. Thus as well as the transverse Alfvén wave propagating along the field there is also a slow compressive wave which travels in the same direction (Friedrichs and Kranzer 1958). This

slow wave has been invoked by Lighthill (1960) to explain the erratic reflection of radar waves from rockets moving through the ionosphere, the slow wave producing an increase in electron density along the earth's magnetic field lines sufficient to produce an additional reflection which depends on the direction of the radar beam relative to the earth's field.

1.5. *Stability*

The theory of the stability of stationary magnetohydrodynamic systems is closely tied to the theory of small oscillations. It, too, starts from the linearized equations (1.22), (1.23) and may proceed in either of two ways. The first procedure is to consider the partial differential equations (Kruskal and Schwarzschild 1954) formed by (1.23) for a given equilibrium configuration $p_0(r)$, $\mathbf{B}_0(r)$ such that

$$\nabla p_0 = (\mathrm{curl}\, \mathbf{B}_0) \times \mathbf{B}_0. \qquad \ldots\ldots(1.28)$$

These equations, together with appropriate boundary conditions, may be used to determine a set of normal modes for $\boldsymbol{\xi}$ and if time dependence is taken as $\exp(i\omega t)$, (1.23) and the boundary conditions form an eigenvalue problem for ω^2. If all the eigenvalues are real and positive, any disturbance about equilibrium will oscillate, so that the system remains near its original configuration. If, however, one or more of the eigenvalues ω^2 is negative, or complex, a growing solution is possible and there exists a disturbance capable of taking the system away from equilibrium. It is often possible to demonstrate that the eigenvalue ω^2 is a continuous function of some parameter characterizing the solution, e.g. of the wave number in a Fourier decomposition, whereupon it is enough to show that the eigenvalue ω^2 passes through zero. Then it is not necessary to solve the eigenvalue problem, but enough to show that the perturbed equilibrium equations have a solution, the marginally stable mode.

The normal mode analysis has found extensive applications, but usually in systems of a high degree of symmetry. One of its successes was the prediction of the ' stabilized pinch ' (Tayler 1957). This is a plasma cylinder which contains within it an axial field and outside a helical field, bounded by a cylindrical perfectly conducting wall, the two field regions being separated by a current sheet on the surface of the plasma. This system is characterized by three parameters, the ratio b of the external axial field B_z^e to the internal axial field B_z^i, β, the ratio of the gas pressure to the internal magnetic pressure $8\pi p/(B_z^i)^2$ and the ratio Λ between the radius of the conducting vessel and the radius of the plasma, and for a suitable choice of these parameters the system is stable.

As a more powerful alternative a variational procedure may be used (Bernstein *et al.* 1958). This depends on the existence of the energy integral (1.10). Because of the conservation of the integrated energy, the linearized equations of motion (1.23) may be written as

$$\rho \ddot{\xi}_i = F_{ij}\, \xi_j, \qquad \ldots\ldots(1.29)$$

where the operator F_{ij} is self-adjoint in the sense that

$$\int \eta_i F_{ij}\, \xi_j \, d\tau = \int \xi_i F_{ij}\, \eta_j \, d\tau, \qquad \ldots\ldots(1.30)$$

where the integral extends over the entire volume of the system, which is assumed to be bounded by rigid walls. This is the case, since the existence of the energy integral implies

$$\frac{d}{dt}\int \tfrac{1}{2}\rho\dot{\xi}^2 = \int \dot{\boldsymbol{\xi}}.\rho\ddot{\boldsymbol{\xi}}\,d\tau = \int \dot{\xi}_i F_{ij}\,\xi_j = \frac{1}{2}\int \frac{d}{dt}(\xi_i F_{ij}\,\xi_j)$$

$$= \frac{1}{2}\int (\dot{\xi}_i F_{ij}\,\xi_j + \xi_i F_{ij}\,\dot{\xi}_j)\,d\tau,$$

where $\dot{\boldsymbol{\xi}}$ and $\boldsymbol{\xi}$ may be arbitrary functions of position. Thus

$$\delta W = -\frac{1}{2}\int d\tau\,\xi_i F_{ij}\,\xi_j \qquad \ldots\ldots(1.31)$$

a bilinear functional of the vector $\boldsymbol{\xi}(r)$ has the role of the potential energy for the linearized equation (1.23). As is usual for conservative systems an equilibrium is stable only if there exists no perturbation $\boldsymbol{\xi}$ which reduces the potential energy, i.e. for which δW is negative. Furthermore, the existence of the energy integral establishes that ω^2 is necessarily real.

Explicitly the plasma contribution to δW may be written

$$\delta W_p = \frac{1}{2}\int [Q^2 - \mathbf{j}_0.\boldsymbol{\mathcal{Q}}\times\boldsymbol{\xi} + \gamma p_0(\operatorname{div}\boldsymbol{\xi})^2 + (\operatorname{div}\boldsymbol{\xi})\,\boldsymbol{\xi}.\nabla p_0]\,d\tau, \quad \ldots\ldots(1.32)$$

where $\boldsymbol{\mathcal{Q}} = \operatorname{curl}(\boldsymbol{\xi}\times\mathbf{B}_0)$ is the perturbed magnetic field.

There is also a contribution $\delta W_v = \int \tfrac{1}{2}B^2\,d\tau$, from the vacuum, and a surface contribution $\int[\![\tfrac{1}{2}B^2 + p]\!]\,ds$, where $[\![\phi]\!]$ is the jump in ϕ at a surface of discontinuity. The only necessary boundary condition is $B_n = 0$ on the perturbed surface, pressure balance being replaced by the surface contribution. The $\boldsymbol{\xi}$ must be normalized, but this need be applied only to the component normal to the magnetic surfaces. Particularly in the hands of the Princeton group the energy method has been extremely powerful. It has been applied to the stability of toroidal configurations, and to a helical plasma, probably the most elaborate problem that has been solved.

There are certain subtleties in the treatment of plasma stability. For example, for a cylinder of plasma, perturbations may be analysed as

$$\boldsymbol{\xi}(r) = \sum_{m,k} \boldsymbol{\xi}_{mk}(r)\exp[i(m\theta + kz)]$$

where r, θ, z are cylindrical polar coordinates, and the Fourier components handled independently. The Euler equations of δW ($\partial\delta W/\partial\xi = 0$) may be formed, are algebraic for ξ_θ, ξ_z, and imply $\operatorname{div}\boldsymbol{\xi} = 0$, except for helical fields of constant pitch. These conditions, which minimize δW with respect to ξ_θ, ξ_z, may be used to eliminate ξ_θ, ξ_z, whereupon δW_p may be written in terms of ξ_r alone as

$$\delta W_p = \int_0^R r\,dr\left\{\xi_r^2\left[\left(\frac{mB_\theta}{r} + kB_z\right)^2 - \frac{2B_\theta}{r}\frac{d}{dr}(rB_\theta) + \left(\frac{mB_\theta}{r} - kB_z\right)^2\Big/(m^2 + k^2r^2)\right]\right.$$

$$\left. + 2r\xi_r\frac{d}{dr}\xi_r\cdot\left(k^2B_z^2 - \frac{m^2B_\theta^2}{r^2}\right)\Big/(m^2 + k^2r^2) + \left(r\frac{d\xi_r}{dr}\right)^2\left(\frac{mB_\theta}{r} + kB_z\right)^2\Big/(m^2 + k^2r^2)\right\},$$

$$\ldots\ldots(1.33)$$

or, introducing $f = kB_z + \frac{m}{r}B_\theta$; $g = kB_z - \frac{m}{r}B_\theta$; $h = \frac{2B_\theta}{r^2}\frac{d}{dr}(rB_\theta)$,

$$\delta W_p = \int r dr \left\{ \frac{(rf\xi' + g\xi)^2}{m^2 + k^2 r^2} + (f^2 - h)\xi^2 \right\}. \qquad \ldots\ldots(1.34)$$

Note that this can be unstable only if h is positive, i.e. if B_θ falls off less rapidly than $1/r$. In particular if the plasma lies in an annulus and is confined by a current flowing in a rigid axial conductor interacting with reversed current in the plasma, $h < 0$ and the configuration is stable. This is the 'hard-core' or 'unpinch' configuration.

The Euler equation for (1.34) may be written

$$\frac{d}{dr}\left\{ \frac{rf(rf\xi' + g\xi)}{m^2 + k^2 r^2} \right\} - \frac{g(rf\xi' + g\xi)}{m^2 + k^2 r^2} - (f^2 - h)\xi = 0, \qquad \ldots\ldots(1.35)$$

which has a singularity at $f = 0$, i.e. $kB_z(r_0) + (m/r_0)B_\theta(r_0) = 0$. The singularity dominates the solution, and associated with it is a characteristic local instability, whose character may be determined by expanding about the singularity, and depends upon the value of the parameter

$$M^2 = -\left[\frac{8\pi p'}{rB_z^2} \left(\frac{\mu}{\mu'} \right)^2 \right], \quad \mu = \frac{B_\theta}{rB_z}.$$

If $M^2 > \frac{1}{4}$, the solutions have been shown unstable, hence a necessary condition for stability is

$$\frac{r}{4}\left(\frac{\mu'}{\mu} \right)^2 + \frac{8\pi p'}{B_z^2} \geqslant 0, \qquad \ldots\ldots(1.36)$$

the Suydam criterion (Suydam 1958). These local instabilities associated with singular points play an important role in magnetohydrodynamics (Rosenbluth 1958).

If the pitch μ is constant over some region, the role of the local instabilities is taken over by simple convective instabilities, in which flux tubes are interchanged and the energy is derived from the expansion of the gas. These will grow in a cylinder if

$$B_\theta \frac{d}{dr}\left(\frac{B_\theta}{r} \right) + \frac{2B_\theta^2}{r^2(B_\theta^2 + B_z^2 + \gamma_p)} > 0.$$

A necessary and sufficient condition for the stability of a cylinder of plasma in the magnetohydrodynamic approximation has been given by Newcomb (1959). It too requires a consideration of the singular points, and may be stated thus:

A necessary and sufficient condition for the magnetohydrodynamic stability of a plasma cylinder is that the solutions to the Euler equations of the energy integral which are small at the singular points shall have no zeros between these singular points.

This rests upon the following argument. Consider a system bounded by $r = r_1$ and $r = r_5$. If the Euler solution ξ_1 which vanishes at r_1 vanishes again at r_2 ($< r_5$), it contributes nothing to the energy integral between these limits, since

for an Euler solution the integrand of δW becomes a perfect differential and the integral is proportional to the value of ξ at the limits; thus

$$\delta W = \left[\frac{r^2 f \xi(rf\xi' + g\xi)}{m^2 + k^2 r^2}\right]_{r_1}^{r_5} = 0.$$

Since ξ need not have a continuous derivative, a possible perturbation is ξ_1, $r_1 < r < r_2$ and $\xi_2 \equiv 0$, $r_2 < r < r_5$. For this perturbation the total integral becomes zero.

However, it is possible to replace part of this discontinuous perturbation by a second Euler solution ξ_3 for which $\xi_3 = \xi_1$ at a radius r_3 where $r_1 < r_3 < r_2$ and which has a zero at r_4 where $r_2 < r_4 < r_5$. By this means the value of the energy integral is reduced, hence it must become negative and the system is necessarily unstable.

Note that in these demonstrations an essential role is played by discontinuous perturbations, or by perturbations with discontinuous derivatives. Such objects are permissible within the context of *ideal* magnetohydrodynamics which permits discontinuities in the currents and the velocity shears. Now a realistic model of the plasma would not permit such phenomena, but since any realistic plasma model must be more elaborate than magnetohydrodynamics, including more information about the internal energy of the plasma, it is not obvious that stability criteria will become less stringent. However, it is clear that the magnetohydrodynamics stability criteria, while useful, are certainly artificial.

At the same time, there is a modest agreement between theory and experiment in the study of magnetohydrodynamic stability. It should, however, be observed that many types of instability are almost independent of the particular properties assigned to the fluid; for example, the gross wriggling of a constricted discharge, the earliest instability to be noted (Carruthers and Davenport 1957), requires only that the plasma be a flexible conductor, hence it is scarcely surprising that it was found. More recently fast tracing cameras with exposure times of less than 1 μsec have displayed the appearance of Rayleigh–Taylor instability in rapidly collapsing cylindrical systems for axial current (Niblett, private communication), for axial field (Curzon *et al.* 1960), and for helical fields. These occur more or less as predicted by theory, but have a finite wavelength, as might be expected in a real fluid.

Attempts to investigate the detailed consequences of magnetohydrodynamic stability theory have been only moderately successful. Magnetic probe measurements in a rapidly collapsing pinch showed evidence of fluctuations appearing most strongly in regions where the Suydam criterion is violated (Burkhardt and Lovberg 1958). On the other hand, attempts to produce the stable unpinch predicted by (1.34) have not been successful (Birdsall *et al.* 1960, Aitken *et al.* 1960), for there remain strong fluctuations on the magnetic probe traces, which occur somewhat later than in a corresponding pinch. These fluctuations may be attributed to electrode effects, and as the conditions for stability in a toroidal system are unknown but probably stringent, it is not surprising that no completely stable configuration has yet been produced. One of the more promising approaches to such a configuration is the toroidal unpinch or levitron in which the central conductor is allowed to fall freely during the 100 μsec course of the experiment (Aitken *et al.* 1960).

At present the situation rather resembles that in the study of hydrodynamic stability, e.g. of Poisseuille flow, where the idealizations made in the theory are almost impossible to realize in the laboratory (Lin 1955).

§2. The Kinetic Theory of a Dense Plasma

The magnetohydrodynamic description of the plasma has had wide application and scored many successes. However, in view of the lack of an extensive body of experimental confirmation of theoretical predictions, and of the somewhat artificial nature of the assumptions underlying the model, there is often a slight air of unreality about magnetohydrodynamic calculations. The rest of this article will be devoted to the study of plasma models with an apparently more realistic basis. The most satisfactory way to proceed would be to determine the relevant equation of state from experiment, but in default of this it is necessary to have recourse to an underlying theory of the dynamics of the plasma, and to use the methods of kinetic theory to deduce appropriate expressions for the pressure tensor p_{ij} and the electric current $\mathbf{j}$. It should be stressed that only at these points does the physical nature of the fluid enter the magnetohydrodynamic description, since until the pressure and current are prescribed Eqn (1.2) is an empty formal statement.

In some ways the simplest, and certainly the most completely developed, approach to this problem is through the classical kinetic theory of diffuse gases. This has as its starting point Boltzmann's equation (Chapman and Cowling 1951) for the development of the distribution function $f(\mathbf{r}, \mathbf{v}, t)$,

$$\frac{\partial f}{\partial t}+\mathbf{v}.\frac{\partial f}{\partial \mathbf{r}}+\mathbf{a}.\frac{\partial f}{\partial \mathbf{v}}=-\int d^3\mathbf{v}'\sigma(\mathbf{v}-\mathbf{v}')|\mathbf{v}-\mathbf{v}'|[f(\mathbf{v})f(\mathbf{v}')-f(\bar{\mathbf{v}})f(\bar{\mathbf{v}}')], \quad \ldots\ldots(2.1)$$

an equation which we will frequently abbreviate as

$$Df = I(f,f). \quad \ldots\ldots(2.2)$$

In this equation $\mathbf{a}(\mathbf{r}, \mathbf{v}, t)$ is considered not as the actual acceleration experienced by a particle, but as the acceleration due to macroscopic forces, which may depend on f, but not on the details of the actual configuration. An alternative and more fundamental starting point is Liouville's equation (Fowler 1936) for the distribution function $f_N(\mathbf{r}_1 \ldots \mathbf{r}_N, \mathbf{v}_1 \ldots \mathbf{v}_N, t)$ of the entire system.

$$\frac{\partial f_N}{\partial t}+\sum_i \mathbf{v}_i.\frac{\partial f_N}{\partial \mathbf{r}_i}+\sum_i \mathbf{a}_i(\mathbf{r}_1 \ldots \mathbf{r}_N, \mathbf{v}_1 \ldots \mathbf{v}_N, t).\frac{\partial f_N}{\partial \mathbf{v}_i}=0, \quad \ldots\ldots(2.3)$$

an equation which for conservative systems is exactly equivalent to Newton's laws of motion, and whose solution is equivalent to determining exactly the motions of all the particles making up the system. This may be reduced to the Boltzmann equation by first integrating over the coordinates of all but one of the particles, reducing f_N to f, in all except the last term. Here, since the acceleration includes terms representing particle interactions, we are left with

$$\sum_j \int \mathbf{a}_i(\mathbf{r}_i, \mathbf{r}_j).\frac{\partial f}{\partial \mathbf{v}_i}(\mathbf{r}_i, \mathbf{r}_j)\, d^3\mathbf{r}_j\, d^3\mathbf{v}_j,$$

which involves the correlation between the positions and velocities of pairs of particles (i,j). The Boltzmann equation is deduced by assuming that particle

interactions occur when a pair of particles are extremely close together, so that the volume in which interactions occur is an infinitesimal fraction of the total volume. This implies that there are never more than two particles within the interaction volume, and that the correlations are negligible except during a collision, and are included in a cross section calculation. It follows that the Boltzmann equation is applicable to diffuse gases in which the molecules interact with short range forces. In the plasma, on the other hand, the forces are long range, and the motion of any particle is always correlated with that of many of its neighbours ; thus the starting point for classical kinetic theory must be re-examined. As we shall show in a later section, (2.1) does form a reasonably good description of the development of the distribution function, provided that the interaction between the particles is modified to allow for charge screening, the principal effect of many-particle correlations. The arguments establishing this involve not the short range, but the relative weakness, of the inter-particle forces, and we will defer them and study not the basis of (2.1) but its consequences.

A prescription was given by Hilbert (1912) for obtaining a solution of (2.1) which leads to normal hydrodynamics and which may be modified to obtain the magneto-hydrodynamic equations. Hilbert began by observing that hydrodynamic behaviour is to be expected when mean free paths are short compared with the distances over which significant changes occur in the distribution function, and that times for changes to occur are large compared to the time between collisions ; hence it is reasonable to look for solutions expanded in powers of the mean free time τ. The consequences of this are easily seen if (2.1) is reduced to dimensionless form, by introducing on the left-hand side a macroscopic scale of length L, and a characteristic velocity V—some average of the particle speeds, and a characteristic time $T = L/V$; then the left-hand side becomes $(1/T)\,Df$, where Df is now in dimensionless form. To treat the right-hand side we introduce a mean density n_0 and a cross section σ_0, so that the right-hand side becomes, with $1/\tau = n_0 \sigma_0 V$,

$$n\sigma V I(f,f) = \frac{1}{\tau} I(f,f),$$

where $I(f,f)$ is dimensionless. Eqn (2.2) may now be written in a form suitable for expansion,

$$I(f,f) = \frac{\tau}{T} Df = \epsilon Df \qquad \ldots\ldots(2.3)$$

and

$$f = f_0 + \epsilon f_1 + \epsilon^2 f_2 + \ldots.$$

When the expansion is substituted in (2.3) and the coefficients of various powers of ϵ equated to zero, a series of integral equations is obtained for the successive approximations to f, of which only the first is non-linear. This first equation is, however, just the equation for equilibrium, hence is satisfied by the locally Maxwellian distribution,

$$f_0 = n(\mathbf{x}, t)\left(\frac{m}{2\pi k T(\mathbf{x})}\right)^{3/2} \exp\left\{-\tfrac{1}{2} m \frac{[\mathbf{v} - \mathbf{u}(\mathbf{x}, t)]^2}{k T(\mathbf{x}, t)}\right\}, \qquad \ldots\ldots(2.4)$$

which contains the five arbitrary functions of space and time, n, $\mathbf{u}$, T.

On introducing $f_i = f_0 \phi_i$, the integral equation for ϕ_1 becomes

$$Df_0 = \int d^3v' K(\mathbf{v}, \mathbf{v}') \phi(\mathbf{v}'), \qquad \ldots\ldots(2.5)$$

where $K(\mathbf{v}, \mathbf{v}')$ is *symmetric*, a consequence of conservation of energy and momentum. Now an equation of this form, for well-behaved K, is consistent only if the left-hand side satisfies a rather stringent condition, namely it must be orthogonal to the solution of the homogeneous equation, i.e. if $\int d^3v' K(\mathbf{v}, \mathbf{v}') h(\mathbf{v}') = 0$, then for a solution Df_0 must satisfy $\int Df_0(\mathbf{v}) h(\mathbf{v}) d^3v = 0$, for :

$$\int Df_0(\mathbf{v}) h(\mathbf{v}) d^3v = \int d^3v \int d^3v' K(\mathbf{v}, \mathbf{v}') \phi(\mathbf{v}') h(\mathbf{v}) = \int d^3v' \phi(\mathbf{v}') \int d^3v K(\mathbf{v}, \mathbf{v}') h(\mathbf{v}) = 0.$$

Thus we obtain a set of integrability conditions on Df_0,

$$\int Df_0 h(\mathbf{v}) = 0.$$

However, the right-hand side of (2.5) is simply the rate of change in ϕ produced by collisions, and, from the conservation laws, there are five invariants m, $m\mathbf{u}$, $\frac{1}{2}mu^2$, and the five associated integrability conditions

$$\left.\begin{aligned} \frac{D\rho}{Dt} + \rho \operatorname{div} \mathbf{u} &= 0, \\ \rho \frac{D\mathbf{u}}{Dt} + \nabla . \mathbf{p} &= 0, \\ \tfrac{3}{2} nk \frac{DT}{Dt} + \mathbf{p} : \nabla \mathbf{u} - \nabla . \mathbf{q} &= 0, \end{aligned}\right\} \qquad \ldots\ldots(2.6)$$

where if

$$\langle \psi \rangle = \frac{\int \psi f d^3v}{\int f d^3v} = \frac{\int \psi f d^3v}{n},$$

$$\left.\begin{aligned} &\rho = nm, \quad \mathbf{u} = \langle \mathbf{v} \rangle \quad \text{and if} \quad \mathbf{c} = \mathbf{v} - \mathbf{u}, \\ &\mathbf{p} = p_{ij} = nm \langle c_i c_j \rangle, \quad \tfrac{3}{2} kT = \tfrac{1}{2} m \langle c^2 \rangle, \\ &\mathbf{q} = n \langle \tfrac{1}{2} mc^2 \mathbf{c} \rangle. \end{aligned}\right\} \qquad \ldots\ldots(2.7)$$

These are the hydrodynamic equations and the associated kinetic theory definitions of the density, mass velocity, pressure tensor, temperature and heat flux vector. The set (2.6) may be used to determine the parameters n, $\mathbf{u}$, T in f_0. At each stage of the approximation procedure a similar constraint is found, giving new values for these quantities. Thus Hilbert's procedure in principle completely determines a particular class of solutions of Boltzmann's equation, the *normal* solutions.

To ensure that the constraints (2.6) are satisfied, the time derivatives of $n, \mathbf{u}, T$ in (2.5) may be eliminated by (2.6), whereupon the integral equation to be solved becomes, with $w^2 = \frac{1}{2}(mc^2/kT)$, $\mathbf{g} = \mathbf{c} - \mathbf{c}'$,

$$f_0(\mathbf{v})\left\{(w_i w_j - \tfrac{1}{3}w^2\,\delta_{ij})\frac{\partial u_i}{\partial x_j} + (\tfrac{5}{2} - w^2)\,\mathbf{c}.\nabla \log T\right\}$$

$$= \int d^3v' \int \sin\theta\, d\theta\, d\phi\, f_0(\mathbf{c}) f_0(\mathbf{c}')\, \sigma(g,\theta)\, g[\Phi(\mathbf{c}) + \Phi(\mathbf{c}') - \Phi(\bar{\mathbf{c}}) - \Phi(\bar{\mathbf{c}}')] \qquad \ldots\ldots(2.8)$$

where $\bar{\mathbf{c}}$ and $\bar{\mathbf{c}}'$ are obtained from $\mathbf{c}$ and $\mathbf{c}'$ by rotating the relative velocity $\mathbf{g} = \mathbf{c} - \mathbf{c}'$, through the scattering angles θ, ϕ.

The technique for solving this begins (Chapman and Cowling 1951) by observing that

$$\Phi = A(w^2)(w_i w_j - \tfrac{1}{3}w^2\,\delta_{ij})\frac{\partial u_i}{\partial x_j} + B(w^2)\, w_i \frac{\partial}{\partial x_i}\log T$$

and determining approximations to the scalar functions A and B. These are usually expanded in Sonine polynomials, $\mathrm{S}_n^{\,m}(x^2)$ defined by

$$\mathrm{S}_n^{\,m}(x^2) = \sum_j \frac{(-1)^j (m+n)!}{(n+j)!\,(m-j)!\,j!}\, x^{2j},$$

which are orthogonal with respect to the kernel $x^{2n+1}\exp(-x^2)$,

$$\int_0^\infty dx\, x^{2n+1}\exp(-x^2)\,\mathrm{S}_n^{\,m}\,\mathrm{S}_n^{\,m'} = \frac{1}{2}\frac{(n+m)!}{n!}\,\delta_{mm'},$$

and a set of algebraic equations is obtained for the coefficients. Alternatively these can be determined from a variational principle (Hirschfelder, Curtis and Bird 1954).

For an ionized gas in a magnetic field the problem although considerably complicated may be tackled in a similar way. The Boltzmann equation becomes

$$\frac{\partial f^{\pm}}{\partial t} + \mathbf{v}.\nabla f^{\pm} + \left(\frac{e}{m}\right)_{\pm}(\mathbf{E} + \mathbf{v}\times\mathbf{B}).\frac{\partial f^{\pm}}{\partial \mathbf{v}} = I(f^{\pm}, f^{\pm}) + I(f^{\pm}, f^{\mp}),$$

a pair of coupled equations. A corresponding complexity is introduced in the hydrodynamic equations. It is possible to form separate moments for the two distributions, and to describe the macroscopic phenomena by the electron density, velocity and pressure, $\rho_-, \mathbf{u}_-, \mathbf{p}_-$, and the corresponding quantities $\rho_+, \mathbf{u}_+, \mathbf{p}_+$, for the ions, thus describing the plasma as two inter-penetrating oppositely charged fluids. For most purposes this is an unsatisfactory procedure since the fluids are closely coupled. If, for example, n_+ and n_- the ion and electron densities differ significantly, large electric fields appear dragging electrons into the positively charged region until quasi-neutrality is established. As we shall see this is an extremely rapid process with a characteristic frequency $\omega_p = (4\pi ne^2/m)^{1/2} \simeq 1000$ Mc/s; hence for all motions except those of very high frequency the plasma can be assumed quasi-neutral. Moreover, if the mean free path is short, there will be frictional forces keeping the difference $(\mathbf{u}_+ - \mathbf{u}_-)$ small, and it is more useful to form moments for the sum of the two equations.

Thus,
$$\int d^3v m_+ \left\{\frac{\partial}{\partial t}+\mathbf{v}.\nabla+\left(\frac{e}{m}\right)_+(\mathbf{E}+\mathbf{v}\times\mathbf{B}).\frac{\partial}{\partial \mathbf{v}}\right\}f^+ + \int d^3v m_- \left\{\frac{\partial}{\partial t}+\mathbf{v}.\nabla+\left(\frac{e}{m}\right)_-(\mathbf{E}+\mathbf{v}\times\mathbf{B}).\frac{\partial}{\partial \mathbf{v}}\right\}f^- = 0$$

or
$$\frac{D\rho}{Dt}+\rho\,\mathrm{div}\,\mathbf{u}=0,$$

where
$$\rho=\rho_+ +\rho_- \quad \text{and} \quad \rho\mathbf{u}=\rho_+\mathbf{u}_+ +\rho_-\mathbf{u}_-.$$

The remaining hydrodynamic equations become

$$\rho\frac{D\mathbf{u}}{Dt}=-\nabla.\mathbf{p}+Q\mathbf{E}+\mathbf{J}\times\mathbf{B}, \qquad \ldots\ldots(2.9)$$

where, if $\mathbf{c}=\mathbf{v}-\mathbf{u}$, note *not* $\mathbf{v}-\mathbf{u}_-$, $\mathbf{v}-\mathbf{u}_+$,

$$p_{ij}=\rho_+\langle c_i^+ c_j^+\rangle+\rho_-\langle c_i^- c_j^-\rangle,$$
$$Q=n_+e_+ +n_-e_-,$$
$$\mathbf{J}=n_+e_+\mathbf{u}_+ +n_-e_-\mathbf{u}_- = Q\mathbf{u}+\langle m_+e_+\mathbf{c}_+ +m_-e_-\mathbf{c}_-\rangle = Q\mathbf{u}+\mathbf{j}.$$

With the last transformation (2.9) becomes

$$\rho\frac{D\mathbf{u}}{Dt}=-\nabla.\mathbf{p}+Q(\mathbf{E}+\mathbf{u}\times\mathbf{B})+\mathbf{j}\times\mathbf{B}. \qquad \ldots\ldots(2.10)$$

Finally, the energy transport is given by

$$\frac{D}{Dt}\tfrac{3}{2}nkT=-\tfrac{3}{2}nkT\,\mathrm{div}\,\mathbf{u}-p_{ij}\frac{\partial u_i}{\partial x_j}-\mathrm{div}\,\mathbf{q}+\mathbf{j}.(\mathbf{E}+\mathbf{u}\times\mathbf{B}), \qquad \ldots\ldots(2.11)$$

where
$$\tfrac{3}{2}nkT=\tfrac{1}{2}\rho_+\langle c_+^2\rangle+\tfrac{1}{2}\rho_-\langle c_-^2\rangle$$

and
$$\mathbf{q}=\tfrac{1}{2}(\rho_+\langle c_+^2\mathbf{c}_+\rangle+\rho_-\langle c_-^2\mathbf{c}_-\rangle).$$

Since the random velocities $\mathbf{c}_+, \mathbf{c}_-$ are defined with respect to the mean velocity $\mathbf{u}$ of the fluid, the temperature and pressure $T, \mathbf{p}$, are not simply related to the temperatures and pressures $T_+, T_-, \mathbf{p}_+, \mathbf{p}_-$ of the component fluids.

It is possible to apply the Hilbert procedure to this coupled system, but if this is done, the zero-order solutions are

$$f_0^+ = n\left(\frac{m_+}{2\pi kT}\right)^{3/2}\exp\left\{-\frac{m_+}{2kT}(\mathbf{v}-\mathbf{u})^2\right\},$$

$$f_0^- = n\left(\frac{m_-}{2\pi kT}\right)^{3/2}\exp\left\{-\frac{m_-}{2kT}(\mathbf{v}-\mathbf{u})^2\right\};$$

thus in zero approximation $\langle c_+|0\rangle$, $\langle c_-|0\rangle=0$ and the currents vanish. Hence we obtain not idealized magnetohydrodynamics, but normal hydrodynamics, with the fluid coupled to the field only in higher approximations.

To obtain a procedure which leads to magnetohydrodynamics we may observe that the term in Boltzmann's equation leading to $\mathbf{j}\times\mathbf{B}$, namely $(e/m)\mathbf{c}\times\mathbf{B}.(\partial f/\partial\mathbf{c})$, is not of the same order as the macroscopic frequencies, but as the cyclotron frequency $eB/m\simeq\omega_L$ ($1{\cdot}7\times10^7\,B$ c/s for electrons); hence it is usually large. It seems reasonable to seek an expansion procedure in which not only the collision

frequency $1/\tau$, but the gyrofrequency ω_L is considered large, and Chapman and Cowling (1951) have proposed an expansion valid for $\omega\tau \simeq 1$

$$I(f,f)+\frac{e}{m}\mathbf{c}\times\mathbf{B}.\frac{\partial f}{\partial \mathbf{c}}=\epsilon\left\{\frac{\partial}{\partial t}+\mathbf{v}.\nabla+\frac{e}{m}(\mathbf{E}+\mathbf{u}\times\mathbf{B}).\frac{\partial}{\partial \mathbf{c}}\right\}f. \qquad \ldots\ldots(2.12)$$

The zero-order solutions may be again taken as Maxwellian, but the zero-order hydrodynamic equations now involve a term $(e/m)\mathbf{c}\times\mathbf{B}.(\partial f/\partial \mathbf{c})$ which involves the first-order distribution function, and which produces the $\mathbf{j}\times\mathbf{B}$ term leading to magnetohydrodynamics. Again integrability conditions are demanded by the symmetry of the total collision integrals, i.e.

$$\int d^3v\psi\left[\left\{\frac{\partial}{\partial t}+\mathbf{v}.\nabla+\frac{e}{m}(\mathbf{E}+\mathbf{u}\times\mathbf{B}).\frac{\partial}{\partial \mathbf{c}}\right\}f^{n-1}+\frac{e}{m}\mathbf{c}\times\mathbf{B}.\frac{\partial}{\partial \mathbf{c}}f^n\right]=0,$$

where ψ is a collision invariant, $m, m\mathbf{v}, \frac{1}{2}mv^2$. These subsidiary conditions involve f^n as well as f^{n-1}, hence are rather different in character from those of the normal solution, but again they may be used to eliminate the time-derivatives $\partial n/\partial t$, $\partial \mathbf{u}/\partial t$, $\partial T/\partial t$ of the parameters in f_0. As before, the stress tensor, current and heat flux vector may be expressed in terms of the gradients of density, velocity, temperature, electric and magnetic field strengths. Also, the perturbation in the distribution $\phi, f_1 = f_0\phi$, may be written in terms of the gradients

$$\phi = B^{ij}\nabla_i u_j - \mathbf{A}.\nabla \log T - n\mathbf{C}.\mathbf{d}, \qquad \ldots\ldots(2.13)$$

where

$$\mathbf{d}_- = -\mathbf{d}_+ = \nabla\left(\frac{n_-}{n}\right)+\frac{n_-n_+(m_+-m_-)}{n\rho p}\nabla p-\frac{n_+n_-}{\rho p}(e_-m_+-m_-e_+)\mathbf{E}. \qquad \ldots\ldots(2.14)$$

Now, however, the solutions may involve in addition to the vector $\mathbf{c}$, the vectors $\mathbf{c}\times\mathbf{B}$, $(\mathbf{c}.\mathbf{B})\mathbf{B}$, and in addition to the scalar c^2, the scalars $(\mathbf{c}.\mathbf{B})^2$ and B^2. Thus for example, the vector $\mathbf{A}$ becomes (Marshall 1957, 1958)

$$\mathbf{A} = \mathscr{A}^{\mathrm{I}}\mathbf{c}+\mathscr{A}^{\mathrm{II}}\mathbf{c}\times\mathbf{B}+\mathscr{A}^{\mathrm{III}}(\mathbf{c}.\mathbf{B})\mathbf{B}; \qquad \ldots\ldots(2.15)$$

in the absence of a magnetic field only the first term would occur. Because of these greater complexities the algebra becomes woefully elaborate; however, it proves possible again to introduce a variational procedure, and determine those coefficients in a Sonine expansion which are responsible for transport effects. Forms such as (2.15) imply that no simple relation exists between the forces and the fluxes, and that besides being tensors, the transport coefficients couple forces and fluxes of different kinds; for example, an electric current may be produced by a temperature gradient as well as by an electric field.

The final forms of the current and heat flux vector are (Marshall 1957, 1958)

$$\mathbf{j} = \sigma^{\mathrm{I}}\mathbf{D}_{\parallel}+\sigma^{\mathrm{II}}\mathbf{D}_{\perp}+\sigma^{\mathrm{III}}\mathbf{h}\times\mathbf{D}+\phi^{\mathrm{I}}(\nabla T)_{\perp}+\phi^{\mathrm{II}}(\nabla T)^{\perp}+\phi^{\mathrm{III}}\mathbf{h}\times\nabla T$$

$$\mathbf{q} = -\lambda^{\mathrm{I}}(\nabla T)_{\parallel}-\lambda^{\mathrm{II}}(\nabla T)_{\perp}-\lambda^{\mathrm{III}}\mathbf{h}\times\nabla T+\psi^{\mathrm{I}}\mathbf{j}_{\parallel}+\psi^{\mathrm{II}}\mathbf{j}_{\perp}+\psi^{\mathrm{III}}\mathbf{h}\times\mathbf{j},$$

where $\mathbf{D}$ is the generalized electric field

$$\mathbf{D} = \mathbf{E}+\mathbf{u}\times\mathbf{B}+\frac{1}{ne}\nabla p$$

and the subscripts $_{\parallel}$, $_{\perp}$ represent components along and across the magnetic field, e.g. $\mathbf{D}_{\parallel} = (\mathbf{D}.\mathbf{h})\mathbf{h}$, $\mathbf{D}_{\perp} = \mathbf{D}-\mathbf{D}_{\parallel}$, where $\mathbf{h}$ is the unit vector along the magnetic field $\mathbf{B}$.

The coefficients may be expressed in terms of the electron collision frequency

$$\tau = \frac{3}{\sqrt{(2\pi)}} \frac{\sqrt{m}(kT)^{3/2}}{ne^4 \log \Lambda}$$

and the electron gyrofrequency

$$\omega = -\frac{eB}{m_-} \simeq 1{\cdot}7 \times 10^7 B \sec^{-1}.$$

The quantity Λ appearing in τ represents the screening effect of distant particles which must be included if the collision integrals are to converge. The correct value, as will be shown, is $\Lambda = (kT/e^2 m^{1/2})^{3/2}$. In terms of these quantities the transport coefficients are found to be (Marshall 1957, 1958, Landshoff 1949, Vaughan-Williams and Haas 1961):

$$\sigma^{\mathrm{I}} = \frac{ne^2\tau}{2m_-} 1{\cdot}931 \qquad = 1{\cdot}931\,\sigma_0$$

$$\sigma^{\mathrm{II}} = \frac{ne^2\tau}{2m_-} \frac{\omega^2\tau^2 + 1{\cdot}802}{\omega^4\tau^4 + 6{\cdot}282\,\omega^2\tau^2 + 0{\cdot}933} \qquad \to \sigma_0/\omega^2\tau^2$$

$$\sigma^{\mathrm{III}} = \frac{ne^2\tau}{2m_-} \frac{-\omega\tau(\omega^2\tau^2 + 4{\cdot}382)}{\omega^4\tau^4 + 6{\cdot}282\,\omega^2\tau^2 + 0{\cdot}933} \qquad \to -\sigma_0/\omega\tau$$

$$\phi^{\mathrm{I}} = \frac{kne\tau}{m_-} 0{\cdot}777 \qquad = 0{\cdot}777\,\phi_0$$

$$\phi^{\mathrm{II}} = -\frac{kne\tau}{m_-} 0{\cdot}75 \frac{\omega^2\tau^2 - 0{\cdot}966}{\omega^4\tau^4 + 6{\cdot}282\,\omega^2\tau^2 + 0{\cdot}933} \qquad \to -0{\cdot}75\,\phi_0/\omega^2\tau^2$$

$$\phi^{\mathrm{III}} = -\frac{kne\tau}{m_-} \frac{2{\cdot}150\,\omega\tau}{\omega^4\tau^4 + 6{\cdot}282\,\omega^2\tau^2 + 0{\cdot}933} \qquad \to -2{\cdot}15\,\phi_0/\omega^3\tau^3$$

and

$$\lambda^{\mathrm{I}} = \frac{kn\tau k T}{m_-} 1{\cdot}020 \qquad = 1{\cdot}020\,\lambda_0$$

$$\lambda^{\mathrm{II}} = \frac{kn\tau k T}{m_-} 1{\cdot}25\left[\frac{1{\cdot}866\,\omega^2\tau^2 + 0{\cdot}966}{\omega^4\tau^4 + 6{\cdot}282\,\omega^2\tau^2 + 0{\cdot}933} + \frac{3m_-/m_+ + 0{\cdot}566(m_-/m_+)^{1/2}}{(\omega^2\tau^2 + 9)m_-/m_+ + 3{\cdot}34(m_-/m_+)^{1/2} + 0{\cdot}32}\right] \to 0{\cdot}707\left(\frac{m_+}{m_-}\right)^{1/2} \frac{\lambda_0}{\omega^2\tau^2}$$

$$\lambda^{\mathrm{III}} = \frac{kn\tau k T}{m_-} 1{\cdot}25\left[\frac{-\omega\tau(\omega^2\tau^2 - 1{\cdot}9)}{\omega^4\tau^4 + 6{\cdot}282\,\omega^2\tau^2 + 0{\cdot}933} + \frac{(m_-/m_+)\,\omega\tau}{(\omega^2\tau^2 + 9)m_-/m_+ + 3{\cdot}34(m_-/m_+)^{1/2} + 0{\cdot}32}\right] \to -1{\cdot}25\frac{\lambda_0}{\omega^2\tau^2}\left[3{\cdot}34\left(\frac{m_+}{m_-}\right)^{1/2} + 13{\cdot}3\right]$$

$$\psi^{\mathrm{I}} = -\frac{kT}{e} 3{\cdot}305 \qquad = -3{\cdot}305\,\psi_0$$

$$\psi^{\mathrm{II}} = -\frac{kT}{e} \frac{2{\cdot}5\,\omega^2\tau^2 + 11{\cdot}5}{\omega^2\tau^2 + 3{\cdot}48} \qquad \to -2{\cdot}5\,\psi_0$$

$$\psi^{\mathrm{III}} = \frac{kT}{e} \frac{1{\cdot}5\,\omega\tau}{\omega^2\tau^2 + 3{\cdot}48} \qquad \to 1{\cdot}5\,\psi_0/\omega\tau.$$

The quantity σ^{III} is the Hall conductivity. There is also some interest in the transverse conductivity $\sigma_\perp$ defined in the absence of temperature and pressure gradients by $\mathbf{j}_\perp = \sigma_\perp \mathbf{E}_\perp$, where $\mathbf{E}_\perp$ is an applied field, and the geometric arrangement is such that no Hall current can flow; then an extra field is induced, the Hall potential, and the conductivity becomes $\sigma_\perp = \sigma_0$.

The stress tensor is determined by the ions, and is best expressed in terms of ion–ion collision time $\tau_+ = (5/2\sqrt{2})\sqrt{(m_+/m_-)}\,\tau$ and the ion gyrofrequency $\omega_+ = -(m_-/m_+)\,\omega_-$.

As usual, it depends upon the divergenceless symmetric tensor formed from the velocity gradients

$$S_{ij} = \frac{1}{2}\left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j}\right) - \tfrac{1}{3}\,\mathrm{div}\,\mathbf{u}\,\delta_{ij}.$$

If the magnetic field is taken along Ox (Marshall 1957, 1958)

$$p_{\perp r} = p - 2\mu S_{xx}$$

$$p_{yy} = p - \frac{2\mu}{1+\frac{16}{9}\omega_+^{\,2}\tau_+^{\,2}}\{S_{yy} + \tfrac{8}{9}\omega_+^{\,2}\tau_+^{\,2}(S_{yy}+S_{zz}) + \tfrac{4}{3}\omega_+\tau_+ S_{yz}\}$$
$$\to p - \mu(S_{yy}+S_{zz}) - \tfrac{3}{2}(\mu/\omega_+\tau_+)\,S_{yz}$$

$$p_{zz} = p - \frac{2\mu}{1+\frac{16}{9}\omega_+^{\,2}\tau_+^{\,2}}\{S_{zz} + \tfrac{8}{9}\omega_+^{\,2}\tau_+^{\,2}(S_{zz}+S_{yy}) - \tfrac{4}{3}\omega_+\tau_+ S_{yz}\}$$
$$\to p - \mu(S_{zz}+S_{yy}) + \tfrac{3}{2}(\mu/\omega_+\tau_+)\,S_{yz}$$

$$p_{xy} = p_{yx} = \frac{-2\mu}{1+\frac{4}{9}\omega_+^{\,2}\tau_+^{\,2}}\{S_{xy} + \tfrac{2}{3}\omega_+\tau_+ S_{xz}\}$$
$$\to (-3\mu/\omega_+\tau_+)\,S_{xz}$$

$$p_{xz} = p_{zx} = \frac{-2\mu}{1+\frac{4}{9}\omega_+^{\,2}\tau_+^{\,2}}\{S_{xz} - \tfrac{2}{3}\omega_+\tau_+ S_{xy}\}$$
$$\to (3\mu/\omega_+\tau_+)S_{xy}$$

$$p_{yz} = p_{zy} = \frac{-2\mu}{1+\frac{16}{9}\omega_+^{\,2}\tau_+^{\,2}}\{S_{yz} - \tfrac{2}{3}\omega_+\tau_+(S_{yy}-S_{zz})\}$$
$$\to (3\mu/4\omega_+\tau_+)\,(S_{yy}-S_{zz}),$$

where $$\mu = \tfrac{1}{3}nkT\tau_+.$$

In this solution there are certain terms which persist and remain finite as the mean free time goes to infinity, but in the limit, viscosity does not lead to energy dissipation.

The treatment of the normal solution of the Boltzmann equation has been rather fully developed and the resultant hydrodynamic system has a reasonable though complex form. However, the nature of the approximation must be borne in mind. It is an expression valid when the macroscopic frequencies are much less than the collision frequency. In a magnetic field the approximation requires that the gyrofrequency should be close to the collision frequency, $\omega\tau \simeq 1$, and the coefficients are obtained as rational functions of $\omega\tau$, and as higher approximations are made, higher powers of $\omega\tau$ appear in both numerator and denominator, hence at large $\omega\tau$

some instability may be expected. Furthermore the value of the product $\omega\tau$ can often be large. Table 2 contains a comparison of periods in various plasmas, and it will be seen that while for some astrophysical situations the normal solution approximation is justified, in many laboratory situations, particularly when the plasma is hot, the value of $\omega\tau$ is large, and even the hydrodynamic frequencies can exceed the collision frequency; thus it appears that some new approximation

Table 2. Some Typical Plasmas

	n	T	ω_p	λ_D	$n\lambda_D^3$	τ^{-1}	$1/\omega_p\tau$
Interstellar gas	1	10^4	6.10^4	10^3	10^9	10^{-4}	2.10^{-7}
Gaseous nebulae	10^3	10^4	6.10^6	30	3.10^7	0·1	5.10^{-8}
Solar corona	10^8	10^6	6.10^7	1	10^8	0·1	2.10^{-9}
Solar atmosphere	10^{14}	10^4	6.10^{11}	10^{-4}	10^2	3.10^9	5.10^{-3}
Gas discharge	10^{14}	10^4	6.10^{11}	10^{-4}	10^2	3.10^9	5.10^{-3}
Hot plasmas	10^{14}	10^5	6.10^{11}	3.10^{-4}	3.10^3	10^8	2.10^{-4}
	10^{14}	10^8	6.10^{11}	10^{-2}	10^5	3.10^8	6.10^{-6}
Diffuse plasmas	10^{12}	10^6	6.10^{10}	10^{-2}	10^8	10^5	2.10^{-7}
Dense plasmas	10^{16}	10^6	6.10^{12}	10^{-4}	10^4	5.10^{-6}	5.10^{-5}
Thermonuclear plasmas	10^{16}	10^8	6.10^{12}	10^{-2}	10^7	8.10^5	4.10^{-6}

n is density (cm^{-3}), T, temperature (°K), ω_p, plasma frequency (sec^{-1}), λ_D, Debye length (cm), τ^{-1}, collision frequency (sec^{-1}). $n\lambda_D^3 = (kT/4\pi n^{1/3}e^2)^{3/2}$.

Note also the gyrofrequency $eB/mc = 1{\cdot}6 \times 10^7 B\ sec^{-1}$ for electrons
$= 0{\cdot}86 \times 10^4 B$ for protons

where B is in gauss.

procedure is required. One such method is suggested by the study of the high frequency behaviour of the plasma, a treatment in which collisions are completely omitted. The disadvantage of this procedure is that the zero-order solution is undetermined, but perturbations on given distribution functions can usefully be studied. We shall now explore this approach to the dynamics of a hot diffuse plasma.

§ 3. The Kinetic Theory of the Diffuse Unmagnetized Plasma

3.1. *The Vlasov–Landau Theory of High Frequency Phenomena*

The starting point for the theory of the collisionless plasma is a form of the Boltzmann equation for the distribution function in which particle interactions are omitted entirely (Vlasov 1938, Landau 1946). In view of the divergence in the collision cross section arising from the long range nature of the Coulomb interaction this might appear as a somewhat illogical procedure. If, however, instead of examining the collision integrals which must certainly be reconsidered in something other than the naïve original fashion, we observe that the ratio $e^2 n^{1/3}/kT$ between the mean interparticle potential energy and the mean kinetic energy is small (cf. Table 2), it seems reasonable to treat the interaction between particles as a perturbation on their free motion, and to start by neglecting interactions. There will, however, be forces acting on the charges from macroscopic fields and these fields themselves may be modified by the mean properties of the particle motions,

i.e. the bulk charge and current in the plasma. Neglecting interactions, the equation of transport becomes

$$\frac{\partial f}{\partial t}+\mathbf{v}.\frac{\partial f}{\partial \mathbf{x}}+\frac{e}{m}(\mathbf{E}+\mathbf{v}\times\mathbf{B}).\frac{\partial f}{\partial \mathbf{v}}=0, \qquad \ldots\ldots(3.1)$$

which is Vlasov's equation.

We shall first consider the response of the plasma to a small oscillating field, and do this by finding the small modification f' produced on some given distribution f_0 by the field $\mathbf{E}'$. We shall assume that both the modification to the distribution function f' and the perturbing electric field $\mathbf{E}'$ are small and further that before $\mathbf{E}'$ is switched on there are no fields present. We can then linearize the Vlasov equation writing $f=f_0+f'$, observing that

$$\frac{\partial f_0}{\partial t}+\mathbf{v}.\frac{\partial f_0}{\partial \mathbf{x}}=0$$

and the product $\mathbf{E}'f'$ is negligible. (3.1) may then be linearized as

$$\frac{\partial f'}{\partial t}+\mathbf{v}.\frac{\partial f'}{\partial \mathbf{x}}+\frac{e}{m}\mathbf{E}'(\mathbf{x},t).\frac{\partial f_0}{\partial \mathbf{v}}=0. \qquad \ldots\ldots(3.2)$$

The solution to (3.2) is easily written down, since the Green's function for the complete Vlasov equation is known, i.e. we can solve

$$DG=\frac{\partial G}{\partial t}(\mathbf{x},\mathbf{v},t\,;\,\mathbf{x}_0,\mathbf{v}_0,t_0)+\mathbf{v}.\frac{\partial}{\partial \mathbf{x}}G+\mathbf{F}.\frac{\partial}{\partial \mathbf{v}}G=\delta(\mathbf{x}-\mathbf{x}_0)\,\delta(\mathbf{v}-\mathbf{v}_0)\,\delta(t-t_0),$$

for, since the Vlasov equation describes the motion of independent particles, the solution is

$$G(\mathbf{x},\mathbf{v},t\,;\,\mathbf{x}_0,\mathbf{v}_0,t_0)=\delta[\mathbf{x}-\mathbf{X}(t\,;\,\mathbf{x}_0,\mathbf{v}_0,t_0)]\,\delta[\mathbf{v}-\mathbf{V}(t\,;\,\mathbf{x}_0,\mathbf{v}_0,\,t_0)]\,\theta(t-t_0), \qquad \ldots\ldots(3.3)$$

where

$$\mathbf{V}(t\,;\,\mathbf{x}_0,\mathbf{v}_0,t_0)=\mathbf{v}_0+\int_{t_0}^{t}\mathbf{F}[\mathbf{X}(t'\,;\,\mathbf{x}_0,\mathbf{v}_0,t_0),\mathbf{V}(t'\,;\,\mathbf{x}_0,\mathbf{v}_0,t_0)]\,dt'$$

$$\mathbf{X}(t\,;\,\mathbf{x}_0,\mathbf{v}_0,t_0)=\mathbf{x}_0+\int_{t_0}^{t}\mathbf{V}(t'\,;\,\mathbf{x}_0,\mathbf{v}_0,t_0)\,dt'$$

and

$$\theta(t-t_0)=0, \quad t<t_0,$$
$$=1, \quad t\geqslant t_0.$$

This result, which may be verified by direct substitution, explicitly replaces the problem posed by (3.2) by the problem of determining the motion of the independent particles. Given the Green's function the solution to the inhomogeneous equation $DF=I$ may be written

$$F(\mathbf{x},\mathbf{v},t)=\int G(\mathbf{x},\mathbf{v},t\,;\,\mathbf{x}',\mathbf{v}',t')\,I(\mathbf{x}',\mathbf{v}',t')\,d^3\mathbf{x}'\,d^3\mathbf{v}'\,dt'.$$

In (3.2), $\mathbf{F}=0$, $\mathbf{V}=\mathbf{v}_0$, $\mathbf{X}=\mathbf{x}_0+\mathbf{v}_0(t-t_0)$ and

$$f'=-\frac{e}{m}\int_{-\infty}^{t}\mathbf{E}(\mathbf{x}-\mathbf{v}t',t').\frac{\partial f_0}{\partial \mathbf{v}}\,dt'.$$

If the initial distribution f_0 is uniform in space and time, the field $\mathbf{E}$ may be Fourier analysed in space and time, i.e.

$$\mathbf{E}(\mathbf{x}, t) = \sum_{(\mathbf{k},\omega)} \mathbf{E}(\mathbf{k}, \omega) \exp(i\mathbf{k}.\mathbf{x} + i\omega t),$$

whereupon the Fourier components of f' become

$$f' = -\frac{e}{m} \frac{\mathbf{E}(\mathbf{k}, \omega)}{i(\omega + \mathbf{k}.\mathbf{v})} . \frac{\partial f_0}{\partial \mathbf{v}}, \qquad \ldots\ldots(3.4)$$

a result which could have been obtained by inspection from (3.2). The charge and current induced by the field may now be written

$$\left.\begin{aligned} q &= -\frac{e^2}{m} \int \frac{\mathbf{E}(\mathbf{k}, \omega)}{i(\omega + \mathbf{k}.\mathbf{v})} . \frac{\partial f_0}{\partial \mathbf{v}} d^3 v, \\ \mathbf{j} &= -\frac{e^2}{m} \int \mathbf{v} \frac{\mathbf{E}(\mathbf{k}, \omega)}{i(\omega + \mathbf{k}.\mathbf{v})} . \frac{\partial f_0}{\partial \mathbf{v}} d^3 v. \end{aligned}\right\} \qquad \ldots\ldots(3.5)$$

Some care is needed in evaluating these integrals which must be correctly defined at the point $\omega = -\mathbf{v}.\mathbf{k}$, a singularity for real $\mathbf{k}, \omega$. Landau (1946) dealt with this by considering an initial value problem, and using not a Fourier transform but a Laplace transform in the time

$$\mathbf{E}(s, \mathbf{k}) = \int_0^\infty e^{-st} \mathbf{E}(t, \mathbf{k})\, dt.$$

His integrals were then defined in the upper half ω plane and had to be analytically continued to and across the real axis, a continuation which defined the integrals. Alternatively we may make a slight but realistic modification to (3.1), introducing a small interaction term, which can be roughly represented by a relaxation time $-(1/\tau)f'$ on the right-hand side, which indicates that a disturbed distribution relaxes towards f_0. This modifies the integrals to

$$\int \frac{g(v)\, d^3 v}{i(\omega + \mathbf{k}.\mathbf{v}) + (1/\tau)} = \int g(v) \frac{-i(\omega + \mathbf{k}.\mathbf{v}) + (1/\tau)}{(\omega + \mathbf{k}.\mathbf{v})^2 + (1/\tau)^2} d^3 v.$$

The Landau result is now obtained by going to the limit $(1/\tau) \to 0$

$$\begin{aligned} &\lim_{(1/\tau)\to 0} \int \frac{dv\, g(v)}{(\omega + \mathbf{k}.\mathbf{v})^2 + (1/\tau)^2} \left[-i(\omega + \mathbf{k}.\mathbf{v}) + \frac{1}{\tau} \right] \\ &= P \int \frac{g(v)}{i(\omega + \mathbf{k}.\mathbf{v})} + \left\{ \frac{1}{k} \lim_{(1/\tau)\to 0} \frac{1}{\tau} \int \frac{g(x)}{x^2 + (1/\tau)^2} dx = \frac{\pi}{k} g(0) \right\} \end{aligned}$$

where P represents the principal part of the integral.

Hence

$$\frac{1}{i(\omega + \mathbf{k}.\mathbf{v})} = P \frac{1}{i(\omega + \mathbf{k}.\mathbf{v})} + \pi \delta(\omega + \mathbf{k}.\mathbf{v}).$$

The δ function contribution here is due to Landau (1946). It leads ultimately to a damping of oscillations which is independent of collision frequency and is the result of a resonance between particle motions and the phase velocity of oscillations. This is known as Landau damping.

If the distribution function f_0 is isotropic in the velocities, i.e. $f_0(v^2)$, (3.5) may be simplified somewhat; for if the electric fields are split into components parallel and perpendicular to the direction $\hat{\mathbf{k}}$ of propagation,

$$\mathbf{E}_\parallel = (\mathbf{E}.\hat{\mathbf{k}})\hat{\mathbf{k}}, \quad \mathbf{E}_\perp = \mathbf{E} - \mathbf{E}_\parallel,$$

$$q = -\frac{2e^2}{m}\int d^3v \frac{E_\parallel v_\parallel + \mathbf{E}_\perp.\mathbf{v}_\perp}{i(\omega + kv_\parallel)}\frac{\partial f_0}{\partial v^2} = -\frac{2e^2 E_\parallel}{m}\int \frac{d^3v\, v_\parallel (\partial f_0/\partial v^2)}{i(\omega + kv_\parallel)},$$

the second integral vanishing since it is odd in the component $v_\perp$. Similarly

$$j_\parallel = -\frac{2e^2 E_\parallel}{m}\int \frac{d^3v\, v_\parallel{}^2}{i(\omega + kv_\parallel)}\frac{\partial f_0}{\partial v^2},$$

$$j_{\perp i} = -\frac{2e^2 E_{\perp i}}{m}\int d^3v \frac{v_{\perp i}{}^2}{i(\omega + kv_\parallel)}\frac{\partial f_0}{\partial v^2}; \qquad \ldots\ldots(3.6)$$

thus although $\mathbf{j}_\perp \parallel \mathbf{E}_\perp$ and $\mathbf{j}_\parallel \parallel \mathbf{E}_\parallel$ we cannot say $\mathbf{j} \parallel \mathbf{E}$ since the two integrals are not equal. The equation of continuity is satisfied, for

$$\left(\operatorname{div}\mathbf{j} + \frac{\partial q}{\partial t}\right)_{\mathbf{k},\omega} = i\mathbf{k}.\mathbf{j} + i\omega q = ikj_\parallel + i\omega q$$

$$= -\frac{2e^2 E_\parallel}{m}\int d^3v\, v_\parallel \frac{\partial f_0}{\partial v^2} = 0.$$

We may now discuss the dielectric properties of the plasma. In the low temperature limit dielectric behaviour may be discussed by considering that the displacement of charge is due solely to the velocity introduced by the electric field, and this is obtained by solving the equation of motion for a harmonic field $\mathbf{E}(\omega)e^{i\omega t}$. Then

$$\dot{\mathbf{r}} = \frac{e\mathbf{E}}{m}, \quad \dot{\mathbf{r}}(\omega) = \frac{e\mathbf{E}}{i\omega m}$$

and

$$\mathbf{j}(\omega) = ne\mathbf{v} = ne\dot{\mathbf{r}}(\omega) = \frac{ne^2\mathbf{E}}{i\omega m}.$$

Finally the dielectric coefficient ϵ may be defined using the induced current j_{ind},

$$\frac{\partial \mathbf{D}}{\partial t} = \frac{\partial}{\partial t}(\epsilon\mathbf{E}) = i\omega\epsilon(\omega)\,\mathbf{E}(\omega) = i\omega\mathbf{E} + 4\pi\mathbf{j}_{\text{ind}}(\omega);$$

hence

$$\epsilon_0 = 1 + \frac{4\pi\mathbf{j}.\mathbf{E}}{i\omega E^2} = 1 - \frac{4\pi ne^2}{m\omega^2} = 1 - \frac{\omega_p{}^2}{\omega^2}, \qquad \ldots\ldots(3.7)$$

where ω_p is the plasma frequency. For finite temperatures we may use our solution to Vlasov's equation and write

$$\epsilon_\parallel = 1 - \frac{4\pi e^2}{i\omega m}\int \frac{v_\parallel{}^2}{i(\omega + kv_\parallel)}\frac{\partial f_0}{\partial(\frac{1}{2}v^2)}\,dv,$$

$$\epsilon_\perp = 1 - \frac{4\pi e^2}{i\omega m}\int \frac{v_\perp{}^2}{i(\omega + kv_\parallel)}\frac{\partial f_0}{\partial(\frac{1}{2}v^2)}\,dv,$$

or
$$\epsilon_{\parallel} = 1 - \frac{\omega_p^2}{\omega^2}\left(\frac{-\omega}{k}\right)\int \frac{v_{\parallel}^2}{(\omega/k + v_{\parallel})}\frac{\partial g_0}{\partial(\frac{1}{2}v^2)}\,dv = 1 - \frac{\omega_p^2}{\omega^2}\Phi_{\parallel}\left(\frac{-\omega}{k}\right),$$

$$\epsilon_{\perp} = 1 - \frac{\omega_p^2}{\omega^2}\left(\frac{-\omega}{k}\right)\int \frac{v_{\perp}^2}{(\omega/k + v_{\parallel})}\frac{\partial g_0}{\partial(\frac{1}{2}v^2)}\,dv = 1 - \frac{\omega_p^2}{\omega^2}\Phi_{\perp}\left(\frac{-\omega}{k}\right), \quad \ldots\ldots(3.8)$$

where $f_0 = ng_0$. For a Maxwellian distribution function

$$g_0 = \left(\frac{m}{2\pi kT}\right)^{3/2}\exp\left\{\frac{-mv^2}{2kT}\right\} = \left(\frac{1}{\pi}\right)^{3/2}\frac{1}{v_\theta^3}\exp\left(-v^2/v_\theta^2\right),$$

the functions Φ depend only on the ratios $-\omega/kv_\theta = v_p/v_\theta$ and become

$$\Phi_{\perp} = \frac{x}{\sqrt{\pi}}\int\frac{\exp(-t^2)}{t-x}\,dt = 2x\exp(-x^2)\int_0^x \exp(y^2)\,dy + i\sqrt{(\pi)}\,x\exp(-x^2),$$

$$\Phi_{\parallel} = \frac{2x}{\sqrt{\pi}}\int\frac{t^2\exp(-t^2)}{t-x}\,dt = 2x^2(\Phi_{\perp} - 1). \quad \ldots\ldots(3.9)$$

For small x $\quad \Phi_{\perp} \to 2x^2 + i\sqrt{(\pi)}\,x, \quad \Phi_{\parallel} \to -2x^2 + 2i\sqrt{(\pi)}\,x^3,$

and for large x $\quad \Phi_{\perp} \to 1 + 1/2x^2, \quad \Phi_{\parallel} \to 1 + 3/2x^2.$

Thus for large ω/k, i.e. small k and long wavelength, the low temperature result is recovered.

3.2. *Dielectric Behaviour and the Screened Field*

As usual, the dielectric coefficient may be used to discuss the transmission of radiation through a plasma. But since the possible transmitted frequencies stop at $\omega \simeq \omega_p$, for which the wave number k needed to make $v_p/v_\theta \simeq 1$ is $k \simeq \omega_p/v_\theta \simeq k_D$, the Debye length, which is microscopic, most transmission problems need only the simple, low temperature version of the dielectric coefficient. Note, however, that for long wavelength the dielectric coefficient vanishes for $\omega^2 = \omega_p^2$, hence the plasma can sustain oscillations at the plasma frequency. If the wavelength is finite the oscillations are damped as $\omega_p\sqrt{\pi}(\lambda/\lambda_D)^3\exp\{-(\lambda/\lambda_D)^2\}$.

A more interesting application is to calculate the field produced by a charge moving through a plasma. In an unmagnetized plasma the longitudinal and transverse coefficients separate and we can calculate the longitudinal field from the appropriate form of Poisson's equation

$$\operatorname{div}\epsilon\mathbf{E} = -\operatorname{div}\epsilon\nabla V = 4\pi q(\mathbf{x}, t) = 4\pi q_0\,\delta(\mathbf{x} - \mathbf{v}t),$$

or using a Fourier transformation

$$k^2\,\epsilon V(\mathbf{k}, \omega) = 4\pi q_0\,\delta(\omega + \mathbf{k}.\mathbf{v})$$

and
$$V(\mathbf{k}, \omega) = \frac{4\pi q_0\,\delta(\omega + \mathbf{k}.\mathbf{v})}{k^2\,\epsilon(\mathbf{k}, \omega)} \quad \ldots\ldots(3.10)$$

and
$$V(\mathbf{x}, t) = \frac{4\pi q_0}{(2\pi)^4}\int d^3k\,d\omega\,\frac{\exp[-i(\mathbf{k}.\mathbf{x} + \omega t)]}{k^2\,\epsilon(\mathbf{k}, \omega)}\,\delta(\omega + \mathbf{k}.\mathbf{v}).$$

We may write $k^2\epsilon = k^2 - k_D^2(1/x^2)\phi_1(x)$ and as $x \to 0$ this tends to $k^2 + 2k_D^2$; thus for a charge at rest the potential is

$$V(\mathbf{x}, t) = \frac{q_0}{2\pi^2}\int d^3k \frac{\exp(-i\mathbf{k}.\mathbf{x})}{k^2 + 2k_D^2} = \frac{f_0 \exp(-\sqrt{2}\, k_D r)}{r} \qquad \ldots\ldots(3.11)$$

which represents the screened Coulomb field (Bohm and Pines 1952). If on the other hand the particle is travelling rapidly, $-\omega/k = -v\cos\theta$ is large over most of the range of integration and $k^2\epsilon \simeq k^2(1-\omega_p^2/\omega^2)$; hence the resonance appears at the plasma frequency, and instead of producing a steady solution the potential takes the form of an outgoing cylindrical wave. Thus only if particles are travelling slowly do they produce a static screened Coulomb field. Rapidly moving particles radiate plasma oscillations.

Regardless of the form of the field produced, however, the particle will experience a retarding force (Bohm and Pines 1952) given by the product of the charge and the parallel component of induced electric field evaluated at the origin, i.e.

$$\begin{aligned} F &= \frac{1}{2}\left\{\frac{-4\pi}{(2\pi)^4} q_0^2 \int \frac{i\mathbf{k}.\mathbf{v}}{|v|} \frac{\delta(\omega + \mathbf{k}.\mathbf{v})}{k^2\epsilon(\mathbf{k}, \omega)}\, d^3k\, d\omega + \text{complex conjugate}\right\} \\ &= -\frac{4\pi}{(2\pi)^4} q_0^2 I \int \frac{\mathbf{k}.\mathbf{v}}{|v|} \frac{\delta(\omega + \mathbf{k}.\mathbf{v})}{k^2\epsilon(\mathbf{k}, \omega)}\, d^3k\, d\omega. \qquad \ldots\ldots(3.12) \end{aligned}$$

Writing

$$k^2\epsilon = k^2 - k_D^2\left[X\left(\frac{v_p}{v_\theta}\right) + iY\left(\frac{v_p}{v_\theta}\right)\right]$$

and performing the ω integration yields

$$F = \frac{1}{2\pi^2} q_0^2 \int \frac{d^3k\, k_D^2\, Y(v\cos\theta/v_\theta)}{(k^2 - k_D^2 X)^2 + k_D^4 Y^2} \frac{\mathbf{k}.\mathbf{v}}{|v|}.$$

Since now X and Y depend only on $v\cos\theta/v_\theta$, the volume integral may be written $d^3k = k^2\, dk \sin\theta\, d\theta\, d\phi$, and on carrying out the ϕ integration

$$\begin{aligned} F &= -\frac{1}{\pi} q_0^2 k_D^2 \int d\theta \sin\theta\cos\theta \int dk \frac{k^3\, Y(v\cos\theta/v_\theta)}{(k^2 - k_D^2 X)^2 + k_D^4 Y^2} \\ &= -\frac{1}{4\pi} q_0^2 k_D^2 \int d\theta \sin\theta\cos\theta \left\{ Y \log\left[(k^2 - k_D^2 X)^2 + k_D^4 Y^2\right]\Big|_0^{k_{\max}} \right. \\ &\qquad \left. + \frac{k_D^2 X}{|Y|}\left[\frac{\pi}{2} - \tan^{-1}\frac{X}{|Y|}\right]\right\} \\ &\simeq -\frac{1}{\pi} q_0^2 k_D^2 \int d\theta \sin\theta\cos\theta \left\{ Y \log\left(\frac{k_{\max}}{k_D}\right) + \frac{Y}{4}\log\left[\frac{(1 - k_D^2 X/k_{\max}^2)^2 + (k_D/k_{\max})^4 Y^2}{X^2 + Y^2}\right] \right. \\ &\qquad \left. + \frac{1}{4}\frac{k_D^2 X}{|Y|}\left[\frac{\pi}{2} - \tan^{-1}\frac{X}{|Y|}\right]\right\} \\ &\simeq -\frac{1}{\pi} q_0^2 k_D^2 \frac{v}{v_\theta}\int_0^1 dx\, x^2 \exp\left[-x^2\left(\frac{v}{v_\theta}\right)^2\right] \log\left(\frac{k_{\max}}{k_D}\right). \qquad \ldots\ldots(3.13) \end{aligned}$$

The upper limit $k_{\max}$ to the integration here is determined by the size of the particle if this is large. If, on the other hand, we consider point particles it represents a rather artificial treatment of the correlation between the positions of colliding

particles. In this analysis it is somewhat artificial, but if we consider instead the Boltzmann collision theory of particle interactions the integration over the impact parameter b can be carried out from the origin to some undetermined upper limit $b_{\max}$ and yields a factor $\log(b_{\max}/b_0)$ where $b_0 = (e^2/\frac{1}{2}mv^2)$. We may then identify the upper limit $k_{\max}$ with the reciprocal of the effective minimum impact parameter b_0 and the logarithm in (3.13) approximates to $\log\Lambda$ where $\Lambda^{-1} = (e^2 n^{1/3}/kT)^{3/2}$ must be small if the Vlasov theory is to be applicable. Thus $\log\Lambda$ is large, and the term (3.13) dominates the expression for F. We may compare the frictional coefficient calculated this way with that calculated from the Boltzmann equation for heavy particles moving through an electron gas (Spitzer 1956).

3.3. *The Fokker–Planck Equation and Dielectric Screening*

A second important problem that can be clarified by making use of the dielectric properties of the discharge is the nature of the effective interaction between particles. We can develop an approach to this which differs from Boltzmann collision approximation in that it makes use of the smallness of the interaction potential, and is valid provided that $\Lambda^{-1} = (e^2 n^{1/3}/kT)^{3/2}$ is small. In order to develop the approximation we shall first, following Landau (1937), consider what happens to the Boltzmann integral when the interaction is weak.

If the interaction is weak two things happen; the deflections experienced by the particles become small and the encounters may be treated in the impulse approximation. The Boltzmann integral, in terms of the impact parameter rather than the cross section becomes

$$I(f,f) = \int d^3v' \int b\,db\,d\phi\,g\,[f(\mathfrak{v})f(\mathfrak{v}') - f(\mathbf{v})f(\mathbf{v}')]$$

where $\mathfrak{v}$ and $\mathfrak{v}'$ are those two velocities which after an impact with parameter b will yield final velocities $\mathbf{v}, \mathbf{v}'$ and $\mathbf{g} = \mathbf{v} - \mathbf{v}'$. We now must express $\mathfrak{v}, \mathfrak{v}'$ in terms of $\mathbf{g}$, and the change in $\mathbf{g}$ occurring on a collision. For particles of equal mass the centre-of-mass velocity, which is unchanged on collision, is $\mathbf{V} = \frac{1}{2}(\mathbf{v}+\mathbf{v}')$ while the relative velocity on collision is transformed by a rotation $\mathscr{O}$, depending on b and ϕ, $\mathbf{g} \to \mathscr{O}\mathbf{g}$. Thus the final velocities are determined by

$$\mathbf{V} = \tfrac{1}{2}(\mathfrak{v}+\mathfrak{v}'), \quad \mathscr{O}\mathbf{g} = \mathfrak{v} - \mathfrak{v}' ;$$

hence

$$\mathfrak{v} = \mathbf{V} + \tfrac{1}{2}\mathscr{O}\mathbf{g} = \mathbf{v} - \tfrac{1}{2}(\mathbf{g} - \mathscr{O}\mathbf{g}) = \mathbf{v} - \tfrac{1}{2}\Delta\mathbf{g}$$

where $\Delta\mathbf{g}$ is the change in $\mathbf{g}$ on collision. Similarly $\mathfrak{v}' = \mathbf{v}' + \frac{1}{2}\Delta\mathbf{g}$. Thus

$$I(f,f) = \int d^3v' \int b\,db\,d\phi\,g[f(\mathbf{v}' + \tfrac{1}{2}\Delta\mathbf{g})f(\mathbf{v} - \tfrac{1}{2}\Delta\mathbf{g}) - f(\mathbf{v})f(\mathbf{v}')].$$

Now the approximation that $\Delta\mathbf{g}$ is small may be used to expand the product in powers of $\Delta\mathbf{g}$ and

$$I(f,f) = \int d^3g \int b\,db\,d\phi\,\tfrac{1}{2}g\left[\frac{\partial}{\partial v_i} f(\mathbf{v})f(\mathbf{v}-\mathbf{g}) - f(\mathbf{v})\frac{\partial f}{\partial v}(\mathbf{v}-\mathbf{g})\right]\Delta g_i$$
$$+ \frac{1}{8}\frac{\partial}{\partial v_j}\left[\frac{\partial}{\partial v_i} f(\mathbf{v}).f(\mathbf{v}-\mathbf{g}) - f(\mathbf{v})\frac{\partial f}{\partial v_i}(\mathbf{v}-\mathbf{g})\right]\Delta g_i \Delta g_j. \quad \ldots\ldots(3.14)$$

The ϕ integration abolishes the first terms since $\Delta\mathbf{g}$ must be periodic in the ϕ direction, and in the second terms the impulse approximation may be used to evaluate $\Delta\mathbf{g}$.

$$\Delta\mathbf{g} = \frac{1}{m_r}\int \mathbf{F}_\perp\, dt = \frac{e^2}{m_r}\int_{-\infty}^{\infty} \frac{\mathbf{b}}{(b^2+x^2)^{3/2}}\,\frac{dx}{g} = \frac{e^2}{m_r g}\,\frac{\mathbf{b}}{b^2}. \qquad \ldots\ldots(3.15)$$

It proves convenient to replace the two-dimensional integral d^2b by an integral over three dimensions d^3l, which, since $\mathbf{b}\perp\mathbf{g}$, involves introducing a factor $\delta(\mathbf{l}.\mathbf{g},)$ whereupon

$$\int d^2b\,\Delta g_i\,\Delta g_j = \frac{e^4}{m_r{}^2}\int \frac{1}{g^2}\,\frac{l_i\,l_j}{l^4}\,\delta(\mathbf{l}.\mathbf{g})\,d^3l = \frac{\pi e^4}{m_r{}^2 g^2}\left(\delta_{ij}-\frac{g_i g_j}{g^2}\right)\int\frac{dl}{l}. \qquad \ldots\ldots(3.16)$$

The last integral is now divergent, and this represents our neglect of correlations at large and small b. The latter of these is taken care of in the Boltzmann equation, where it is shown that the lower limit should be $l_0 = 2e^2/m_r g^2$, but the upper limit is undetermined as l_M.

When these limits are substituted,

$$I(f,f) = \frac{\pi e^4}{8m_r{}^2}\log\left(\frac{l_M}{l_0}\right)\left\{-\frac{\partial f(\mathbf{v})}{\partial v_i}\int d^3v'\,\frac{\partial f(\mathbf{v}')}{\partial v_j'}\,\frac{1}{g^3}\,(\delta_{ij}g^2-g_i g_j)\right.$$
$$\left.+\frac{\partial^2 f(\mathbf{v})}{\partial v_i\,\partial v_j}\int d^3v' f(\mathbf{v}')\,\frac{1}{g^3}\,(\delta_{ij}g^2-g_i g_j)\right\}. \qquad \ldots\ldots(3.17)$$

Observe that this has the form

$$\frac{\partial}{\partial v_i}(D_i f)+\frac{\partial^2}{\partial v_i\,\partial v_j}(D_{ij}f),$$

which is that of the Fokker–Planck equation. This description of the time development of the probability distribution for a Markoff process is most easily derived thus: suppose that, given a state characterized by x' at time τ, the probability of finding the state at x at time $t+\tau$ is $w(x-x';\,x',t)$; then given the probability distribution $P(x;\,t)$,

$$P(x;\,t+\tau) = \int P(x';\,t)\,w(x-x';\,x',t)\,dx' = \int P(x-\Delta x,t)\,w(\Delta x;\,x-\Delta x,t)\,d\Delta x$$

and, if w is small enough for any but small Δx, the integrand may be expanded as

$$P(x;\,t)\,w(\Delta x;\,x,t)-\Delta x\,\frac{\partial}{\partial x}\{P(x;\,t)\,w(\Delta x;\,x,t)\}$$
$$+\tfrac{1}{2}(\Delta x)^2\,\frac{\partial^2}{\partial x^2}\{P(x;\,t)\,w(\Delta x;\,x,t)\}+\ldots.$$

But since $\int w\,d\Delta x = 1$, $\int \Delta x w = \langle\Delta x\rangle$; $\int(\Delta x)^2 w = \langle(\Delta x)^2\rangle$, etc., where $\langle\Delta x\rangle$, $\langle(\Delta x)^2\rangle$, etc. are the average of the change in x, the square of the change in x, etc. produced in a transition, then

$$\frac{P(x,t+\tau)-P(x,t)}{\tau} = -\frac{\partial}{\partial x}\left\{\frac{\langle\Delta x\rangle}{\tau}\,P(x,t)\right\}+\frac{1}{2}\,\frac{\partial^2}{\partial x^2}\left\{\frac{\langle(\Delta x)^2\rangle}{\tau}\,P(x,t)\right\}+\ldots$$

and if τ is small enough we obtain the approximate result

$$\left.\begin{aligned} \frac{\partial P}{\partial t} &= -\frac{\partial}{\partial x}(D_1 P)+\frac{\partial^2}{\partial x^2}(D_2 P) \\ \text{where} \qquad D_1 &= \frac{\langle \Delta x \rangle}{\tau}, \quad D_2 = \frac{1}{2}\frac{\langle (\Delta x)^2 \rangle}{\tau}. \end{aligned}\right\} \qquad \ldots\ldots(3.18)$$

It is possible to derive Landau's Fokker–Planck equation for the plasma by considering not the Boltzmann integral, but the fluctuating forces acting on a particle. This may be done by calculating the sum of the Coulomb fields due to its neighbours ; however, unless the long range correlations are taken into account, the long range divergence persists (Thompson and Hubbard 1960).

To consider these correlations we may start from an equation which is certainly correct, provided only that the particle dynamics may be cast in Hamiltonian form. This is Liouville's equation for the complete N-particle distribution function,

$$\frac{\partial}{\partial t} f_N(\mathbf{x}_1 \ldots \mathbf{x}_N, \mathbf{p}_1 \ldots \mathbf{p}_N)+[H(\mathbf{x}_1 \ldots \mathbf{x}_N, \mathbf{p}_1 \ldots \mathbf{p}_N), f_N] = 0, \qquad \ldots\ldots(3.19)$$

where $[X, Y]$ is the Poisson bracket of X, Y ; H is the Hamiltonian for the complete system and f_N the distribution function giving the probability of finding the system in the state characterized by the $6N$ variables $\mathbf{x}, \mathbf{p}$. For non-interacting particles the joint probability distribution f_N may be taken as the product of single-particle functions $f_N = \prod_1^N f_1(\mathbf{x}_i \mathbf{p}_i)$, and (3.19) factorizes into N equivalent Vlasov equations. If particles interact through a potential $\phi(\mathbf{x}_i - \mathbf{x}_j)$, then (3.19) may be integrated over $N-1$ of the coordinates and momenta to give an equation for the one-particle distribution (Green 1952) $f_1(\mathbf{x}_i, \mathbf{p}_i)$,

$$0 = \frac{\partial f_1}{\partial t}+\mathbf{v}_i . \frac{\partial f_1}{\partial \mathbf{x}_i}+\mathbf{F}_i . \frac{\partial f_1}{\partial \mathbf{p}_i}+\sum_j \int d^3 x_j \, d^3 p_j \frac{\partial \phi_{ij}}{\partial \mathbf{x}_i} . \frac{\partial}{\partial \mathbf{p}_i} f_2(\mathbf{x}_i, \mathbf{x}_j, \mathbf{p}_i, \mathbf{p}_j), \qquad \ldots\ldots(3.20)$$

where the last term involves the correlation between the position of pairs of particles. Thus, determining the Boltzmann function for f_1 requires knowledge of the correlation function f_2. An equation for f_2 may be obtained but it involves f_3. To make progress some more or less arbitrary closure assumption must be made relating successive correlation functions—usually expressing f_3 in terms of f_1 and f_2. If the interaction is weak, a fairly systematic procedure has been developed valid in the limit as $e \to 0$, $N \to \infty$, and $Ne^2/m \sim 1$.

The natural appearance of screening in (3.13) where the dielectric coefficient was used in calculating the frictional drag suggests a natural method of treating the effective correlation between particles. We simply calculate the interaction between particles using not the Coulomb interaction $-\nabla^2 \phi = 4\pi q$, but the field calculated as that in a medium with dielectric coefficient ϵ, i.e. from $-\nabla . \epsilon \nabla \phi = 4\pi q$, the dielectric coefficient being calculated from Vlasov's equation. At short distances, k large, and high frequencies, $\epsilon \to 1$ and the interaction is Coulomb ; while at large distances the interaction is modified by many particles and a continuum approximation may be used to represent this modification. In this procedure the correlation of particles does not appear explicitly, but is involved in the modified interaction. Thus the contribution of the particle j to the potential at i is not

simply $\phi_{ij}{}^{C} = e^2/(\mathbf{r}_i - \mathbf{r}_j)$, but includes in addition the contribution $\Sigma_l \delta\phi_{il}$, which represents the modification to the potential at i produced by the particles at l, due to the added potential at l produced by the particle j. In calculating $\delta\phi$, the disturbance in the position of particle l, the motion is linearized, but the disturbance in potential is taken as the modified, not the simple Coulomb potential. This, however, is equivalent to solving the modified Poisson equation.

I shall now demonstrate this procedure by calculating one of the Fokker–Planck diffusion coefficients. It is first useful to express the coefficients in terms of the fluctuating microfield. This is not difficult, for since

$$\dot{\mathbf{v}} = \frac{e}{m}\mathbf{E}(\mathbf{x}, t)$$

$$\Delta\mathbf{v} = \frac{e}{m}\int_{t-\tau}^{t} \mathbf{E}(\mathbf{x}', t')\, dt'$$

and

$$\langle \Delta v_i\, \Delta v_j \rangle = \frac{e^2}{m^2}\left\langle \int_{t-\tau}^{t} dt'\, E_i(\mathbf{x}', t') \int_{t-\tau}^{t} dt''\, E_j(\mathbf{x}'', t'') \right\rangle$$

$$= \frac{e^2}{m^2}\left\langle \int_{t-\tau}^{t} dt'\, E_i(\mathbf{x}', t') \int_{t'-t-\tau}^{t'-t} ds\, E_j[\mathbf{x}(t'-s), t'-s] \right\rangle. \quad \ldots\ldots(3.21)$$

Since the dielectric coefficient is most easily calculated in the Fourier representation $\epsilon(\mathbf{k}, \omega)$, it is useful to express the coefficient in terms of the Fourier components of $\mathbf{E}$. If the field correlations persist for a time much shorter than the time needed for a particle trajectory to be much altered the inner integral may be extended from $-\infty$ to ∞. Characteristic field frequencies are close to ω_p, and particle trajectories change in times of order τ_c, hence the condition is

$$\omega_p \tau_c \gg 1, \quad \text{i.e.} \sqrt{\left(\frac{4\pi n e^2}{m}\right)} \frac{1}{nQv} = \sqrt{\left(\frac{4\pi n e^2}{m}\right)} \frac{\sqrt{m}}{n} \left(\frac{kT}{e^2}\right)^{3/2} \frac{1}{\log \Lambda}$$

$$\simeq \sqrt{(4\pi)} \left(\frac{kT}{e^2 n^{1/3}}\right)^{3/2} \equiv \Lambda^{3/2} \gg 1,$$

which was the condition that the Vlasov equation should be a reasonable approximation. Furthermore, in calculating the quadratic coefficients it is sufficient to consider the trajectories as unperturbed by the field, i.e. $\mathbf{x}(t') = \mathbf{x}(t) - \mathbf{v}(t - t')$, and the coefficients become

$$D_{ij} = \frac{1}{2}\frac{\langle v_i v_j \rangle}{\tau} = \frac{1}{16\pi^3}\frac{e^2}{m^2}\int d^3k \int d\omega \langle E_i E_j{}^*(\mathbf{k}, \omega) + E_i{}^* E_j(\mathbf{k}, \omega) \rangle\, \delta(\omega + \mathbf{k}.\mathbf{v}), \quad \ldots\ldots(3.22)$$

where

$$\mathbf{E}(\mathbf{x}, t) = \frac{1}{(2\pi)^4}\int d^3k \int d\omega\, \mathbf{E}(\mathbf{k}, \omega) \exp\left[i(\mathbf{k}.\mathbf{x} + \omega t)\right] \quad \ldots\ldots(3.23)$$

and $\mathbf{E}^*(\mathbf{k}, \omega) = \mathbf{E}(-\mathbf{k}, -\omega)$ is complex conjugate to $\mathbf{E}(\mathbf{k}, \omega)$.

There remains the problem of determining $\mathbf{E}(\mathbf{k}, \omega)$, which again is fairly simple if we make the approximation suggested above and consider the field due to particles whose trajectories are unmodified by the field, but allow for correlation by using

the dielectric coefficient. Thus if the contribution of the particles to the harmonic components of the charge density is $q(\mathbf{k}, \omega)$ we write

$$\phi(\mathbf{k}, \omega) = \frac{4\pi q(\mathbf{k}, \omega)}{k^2\, \epsilon(\mathbf{k}, \omega)} \qquad \ldots\ldots(3.24)$$

and

$$\langle E_i E_j{}^* \rangle = \frac{16\pi^2 k_i k_j \langle q(\mathbf{k}, \omega)\, q^*(\mathbf{k}, \omega) \rangle}{k^4\, \epsilon(\mathbf{k}, \omega)\, \epsilon^*(\mathbf{k}, \omega)}. \qquad \ldots\ldots(3.25)$$

The q's may be written

$$q(\mathbf{k}, \omega) = \sum_i \exp(i\mathbf{k} \cdot \mathbf{x}_i)\, \delta(\omega + \mathbf{k} \cdot \mathbf{v}_i)$$

and the quantity we require is

$$\left\langle \int d^3k'\, d\omega'\, q(\mathbf{k}, \omega)\, q(\mathbf{k}'\omega') \right\rangle$$
$$= \left\langle \int d^3k'\, d\omega' \sum_i \sum_j \exp(i\mathbf{k} \cdot \mathbf{x}_i) \exp(i\mathbf{k}' \cdot \mathbf{x}_j)\, \delta(\omega - \mathbf{k} \cdot \mathbf{v}_i)\, \delta(\omega' + \mathbf{k}' \cdot \mathbf{v}_j) \right\rangle.$$

Since in this expression there are no correlations, the random phase approximation may be used to reduce the double sum to a single one

$$\left\langle \int d^3k'\, d\omega' \sum_i \exp[i(\mathbf{k} + \mathbf{k}') \cdot \mathbf{x}_i]\, \delta(\omega + \mathbf{k} \cdot \mathbf{v}_i)\, \delta(\omega' + \mathbf{k}' \cdot \mathbf{v}_i) \right\rangle$$

and when we form the phase averages over the $\mathbf{x}_i$, this becomes

$$\int d^3k'\, d\omega'\, d^3v\, f(\mathbf{v})\, \delta(\omega + \mathbf{k} \cdot \mathbf{v})\, \delta(\mathbf{k} + \mathbf{k}')\, \delta(\omega + \omega').$$

Therefore

$$\left\langle \int d^3k'\, d\omega'\, q(\mathbf{k}, \omega)\, q(\mathbf{k}'\omega') \right\rangle = \langle q(\mathbf{k}, \omega)\, q^*(\mathbf{k}, \omega) \rangle = \int d^3v\, f(\mathbf{v})\, \delta(\omega + \mathbf{k} \cdot \mathbf{v}) \qquad \ldots\ldots(3.26)$$

and the diffusion coefficients at v become

$$D_{ij}(\mathbf{v}) = \frac{1}{2} \frac{\langle v_i v_j \rangle}{\tau} = \frac{1}{\pi} \int d^3k\, d\omega\, d^3v' \frac{k_i k_j f(v')\, \delta(\omega + \mathbf{k} \cdot \mathbf{v}')\, \delta(\omega + \mathbf{k} \cdot \mathbf{v})}{k^4\, \epsilon(\mathbf{k}, \omega)\, \epsilon^*(\mathbf{k}, \omega)}. \qquad \ldots\ldots(3.27)$$

For a Maxwellian distribution where $f_0 = n\pi^{-3/2} v_\theta{}^{-3} \exp(-v^2/v_\theta{}^2)$, the only surviving components of D^2 are the diagonal terms $D_\parallel{}^2$ parallel to $\mathbf{v}$, and $D_\perp{}^2$ normal to $\mathbf{v}$

$$D_\parallel{}^2 = \frac{ne^4}{m^2 v_\theta} \frac{1}{\sqrt{\pi}} \int d^3k\, d\omega\, \frac{(\mathbf{k} \cdot \mathbf{v})^2\, \delta(\omega + \mathbf{k} \cdot \mathbf{v}) \exp[-(\omega/kv_\theta)^2]}{v^2[(k^2 - k_D{}^2 X)^2 + k_D{}^4 Y^2(\omega/kv_\theta)]}. \qquad \ldots\ldots(3.28)$$

As in calculating the frictional coefficient, all but the integral over θ in

$$d^3k = k^2\, dk \sin\theta\, d\theta\, d\phi$$

may be carried out, and a term $\log(k_{\max}/k_D)$ appears as dominant, where $k_{\max}$ arises from the *local* short range correlation involved in a collision and D becomes approximately

$$D_\parallel = \frac{8\pi ne^4}{m^2 v} \log \Lambda G\left(\frac{v}{v_\theta}\right), \qquad \ldots\ldots(3.29)$$

where

$$G(x) = \left\{ \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\, dt - x \exp(-x^2) \right\} \Big/ 2x^2,$$

which is exactly the result obtained by Chandrasekhar and by Spitzer (1956). In $D_\perp$, $(\mathbf{k}.\mathbf{v})^2/v^2$ must be replaced by $\{k^2-(\mathbf{k}.\mathbf{v})^2/v^2\}$. In calculating the frictional term $D_\parallel$ it is necessary to use a better approximation to the particle trajectories, writing not $\mathbf{x}+\mathbf{v}t$ but

$$\mathbf{x}+\mathbf{v}t+\int_{t-\tau}^{t}\frac{e}{m}\int^{t'}\mathbf{E}(\mathbf{x}'',t'')\,dt''$$

and expand in the last term; whereupon the surviving term is approximately proportional to $\langle E_i E_j\rangle$. In addition to this the field induced by the test particle itself, while of order $\langle E_t^2\rangle$, $\mathbf{E}_t$ being the fluctuating field, is systematically opposed to the test particle motion. A thorough treatment of this approach to the Fokker–Planck equation has been given, the principal result of which is to verify that (3.17) includes all the dominant terms, provided the dielectric coefficient does not vanish for finite $\mathbf{k}, \omega$. Thus the equivalence of the Boltzmann equation and the Fokker–Planck equation for the dominant terms has been established.

3.4. *Electrostatic Instabilities*

If the dielectric coefficient vanishes for a finite value of ω and $\mathbf{k}$, the method given above for calculating the Fokker–Planck coefficients may break down, the integrals being undefined; however, if so the plasma is unstable, in which case the internal fields may reach values limited only by non-linear phenomena which we have not yet calculated.

To understand this process, as in the magnetohydrodynamic problem, we must first investigate the free propagation of electrostatic oscillations through the system. Since we have already calculated the charge q induced by a field $\mathbf{E}(\mathbf{k},\omega)$, the equation of motion for longitudinal oscillations is

$$-\nabla^2\phi = k^2\phi = 4\pi q_{\text{ind}} = \frac{4\pi e^2}{m}\int d^3v\,\frac{\mathbf{k}.(\partial f_0/\partial \mathbf{v})}{(\omega+\mathbf{k}.\mathbf{v})}\phi,$$

from which is obtained the dispersion relation

$$k^2-\Sigma\frac{4\pi e^2}{m}\int d^3v\,\frac{\mathbf{k}.(\partial f_0/\partial \mathbf{v})}{(\omega+\mathbf{k}.\mathbf{v})}=0, \qquad \ldots\ldots(3.30)$$

where the summation extends over the possible types of particle in the system. Introducing $g=\Sigma(4\pi ne^2/m)\int f_0\,d^2v_\perp$ the dispersion relation becomes

$$k^2=\int dv_\parallel\,\frac{\partial g/\partial v_\parallel}{[(\omega/k)+v_\parallel]}=Z\left(\frac{\omega}{k}\right),$$

where the Landau prescription must be used to define the integral in the lower half plane, and on the real axis. In studying stability we are interested in the behaviour of the complex solutions $\omega(k)$ as a function of real k. For small k

$$Z\left(\frac{\omega}{k}\right)\to\frac{\omega_p^2}{\omega^2}k^2+i\pi\left.\frac{dg}{dv}\right|_{(-\omega/k)},$$

and since the last term is small, $\omega\simeq\pm\omega_p$, and since $g\to 0$ as $v^2\to\infty$, $\mathbf{v}.(\partial g/\partial\mathbf{v})<0$, and the last term always introduces damping. Therefore if instability is to occur, $\omega(k)$ must cross the real axis (Auer 1958, Penrose 1960), i.e. there must exist some real (k,ω) for which $k^2\epsilon(k,\omega)=0$. If g has a single maximum, no instability is

possible. The zero can occur only at this maximum, since $i\pi(\partial g/\partial v)(-\omega/k)$ must vanish, and with no loss in generality the velocity zero may be taken at the maximum, whereupon the real part of the integral becomes $\int(1/v)(\partial g/\partial v)\,dv$ which is negative definite, since $v = 0$ is the sole maximum in g. If the system has more than one maximum an instability is possible, but need not appear, since the principal part must be positive (Buneman 1959, Kharchenko *et al.* 1960).

The simplest and most typical instability occurs when the distribution function represents two well-defined interpenetrating beams (Buneman, Kharchenko *et al.*), whereupon the dispersion relation takes the form

$$1-\frac{\omega_1^2}{(\omega+kv_1)^2}-\frac{\omega_2^2}{(\omega+kv_2)^2}=0, \qquad \ldots\ldots(3.31)$$

where ω_1 and ω_2 are the plasma frequencies of the two streams. As k increases, this quartic has at first four real roots $\omega \simeq 0,\ 0,\ \omega \simeq \pm(\omega_1^2+\omega_2^2)^{1/2}$, but finally two of these become complex and conjugate, one having the form $\omega = \omega_r - i\gamma$. The disturbance corresponding to the root grows as $\exp(i\omega_r t+\gamma t)$, and corresponds to an instability. If $\omega_2^2 \simeq \lambda\omega_1^2$ where λ is small, this occurs for $(kv)^2 = \omega_1^2/(1+\lambda^{1/2})$. For larger values of k the perturbations are unstable. If the two beams are identical, moving with a relative velocity $2u$ the dispersion relation becomes

$$1-\omega_p^2\left\{\frac{1}{(\omega-ku)^2}+\frac{1}{(\omega+ku)^2}\right\}=0 \qquad \ldots\ldots(3.32)$$

with roots

$$\omega^2=\frac{\omega_p^2}{2}\left\{1+2\left(\frac{ku}{\omega_p}\right)^2 \pm \sqrt{\left[1+8\left(\frac{ku}{\omega_p}\right)^2\right]}\right\}$$

which can become negative. Hence $\omega = \pm i\gamma$ for $(ku)^2 \leqslant 2\omega_p^2$.

Two identical beams with a spread in velocities may be represented by the distribution function (Harris 1959)

$$f=\frac{v_\theta}{2\pi}\left[\frac{1}{(v-u)^2+v_\theta^2}+\frac{1}{(v+u)^2+v_\theta^2}\right], \qquad \ldots\ldots(3.33)$$

for which simple function the dispersion relation becomes

$$1-\omega_p^2\left\{\frac{1}{(\omega-ikv_\theta-ku)^2}+\frac{1}{(\omega-ikv_\theta+ku)^2}\right\}=0$$

and the roots are

$$\omega = ikv_\theta \pm \frac{\omega_p}{\sqrt{2}}\left\{1+2\left(\frac{ku}{\omega_p}\right)^2 \pm \sqrt{\left[1+8\left(\frac{ku}{\omega_p}\right)^2\right]}\right\}^{1/2}.$$

For instability ω must be negative and imaginary, and this is possible only if

$$k^2(u^2+v_\theta^2)^2 < \omega_p^2(u^2-v_\theta^2). \qquad \ldots\ldots(3.34)$$

Thus an instability criterion is obtained; the two beams can only be unstable if $u > v_\theta$.

This is a somewhat artificial example for although v_θ may be identified with the velocity spread in the beam, the energy spread with the distribution (3.33) is infinite. Similar analysis may be carried out for Maxwellian distribution functions, but this depends on the properties of the functions (3.9) and must be studied

numerically. It has been shown that for two interpenetrating Maxwellian distributions of ions and electrons the condition on the drift velocity u for the appearance of instability is $u > 0{\cdot}9\, v_\theta$ where $\frac{1}{2}m_-(v_\theta)^2 = kT$.

§4. The Kinetic Theory of the Diffuse Plasma in a Magnetic Field

We are now ready to give a discussion of the properties of a diffuse plasma in a magnetic field, in which much more complicated behaviour can be expected than was found in the absence of such a field. First we will consider the dielectric properties of a homogeneous magnetized plasma, which are quite involved. Instead of exploring these fully, however, we shall go ahead to examine phenomena occurring when the fields are not uniform. Here the fact that particles are localized on magnetic lines of force proves capable of producing a quasi-magnetohydrodynamic behaviour even though collisions are entirely absent. We shall discover that even at hydrodynamic frequencies far above the collision frequency where the normal solution cannot be used the plasma is still capable of displaying hydrodynamic behaviour, at least for motions across magnetic field lines. For motions along the field lines quite different behaviour is to be expected, and a different description must be used.

4.1. *The Dielectric Properties of a Cold Magnetized Plasma*

The first problem we shall consider is that of determining the response of a cold uniform plasma in a uniform magnetic field to an applied electric field (Brown 1960) and obtaining the dielectric coefficient analogous to (3.7). The simplification introduced by considering the plasma as cold is that thermal motions can be neglected. Furthermore, since thermal motion only enters through the combination kv_θ, the resulting description should be valid for wavelengths long compared with the Debye length.

The equation of motion for a particle becomes

$$\dot{\mathbf{v}} = \frac{e}{m}(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \qquad \ldots\ldots(4.1)$$

where $\mathbf{B}$ is a constant uniform magnetic field, and $\mathbf{E}$ is uniform in space but varies sinusoidally in time. The motion is most easily described by introducing the velocity component u along the magnetic field and the complex velocity of rotation about the field line $w = v_y + iv_z$ where $\mathbf{B} \| Ox$. We obtain

$$\begin{aligned}
\dot{u} &= \frac{e}{m}E_x, \\
\dot{w} = \dot{v}_y + i\dot{v}_z &= \frac{e}{m}(E_y - v_z B) + \frac{ie}{m}(E_z + v_y B) \\
&= \frac{e}{m}(E_y + iE_z) + \frac{ieB}{m}(v_y + iv_z) \\
&= \frac{e}{m}\mathscr{E} + i\Omega w. \qquad \ldots\ldots(4.2)
\end{aligned}$$

In the absence of an electric field the motion is described by $w = w_0 e^{-i\Omega t}$ or, writing $w = v_\perp e^{i\phi}$, by $v_\perp$ = constant, $\phi = \phi_0 - \Omega t$, i.e. a rotation about the field line with the gyrofrequency $\Omega = eB/m$. When an electric field of frequency ω is applied the solutions reach a steady state where

$$u = -\frac{ie}{m\omega} E_x$$

and

$$w = -\frac{ie}{m(\omega - \Omega)} \mathscr{E}$$

or in terms of E_x and E_y

$$v_y = \frac{e}{m}\left[-\frac{i\omega E_y}{\omega^2 - \Omega^2} + \frac{\Omega E_z}{\omega^2 - \Omega^2}\right] e^{i\omega t} + \text{complex conjugate}$$

$$v_z = \frac{e}{m}\left[-\frac{\Omega E_y}{\omega^2 - \Omega^2} - \frac{i\omega E_z}{\omega^2 - \Omega^2}\right] e^{i\omega t} + \text{complex conjugate.} \qquad \ldots\ldots(4.3)$$

Because of these motions a current will be set up which is proportional to the magnitude of the electric field, but not parallel to it. In a plasma of density n, and including the effect of the induced motion of the ions (charge e, mass M),

$$\begin{aligned}
j_x &= -\frac{ne^2}{m}\frac{i\omega}{\omega^2}\left(1 + \frac{m}{M}\right) E_x = -\frac{i\omega}{4\pi}\frac{\omega_p^2}{\omega^2}\left(1 + \frac{m}{M}\right) E_x \\
j_y &= -\frac{ne^2}{m} i\omega \left\{\left[\frac{E_y}{\omega^2 - \Omega^2} - \frac{i\Omega E_z}{\omega(\omega^2 - \Omega^2)}\right] + \frac{m}{M}\left[\frac{E_y}{\omega^2 - \Omega_+^2} - \frac{i\Omega_+ E_z}{\omega(\omega^2 - \Omega_+^2)}\right]\right\} \\
&= -\frac{i\omega}{4\pi}\omega_p^2 \left\{E_y\left[\frac{1}{\omega^2 - \Omega^2} + \frac{m}{M}\frac{1}{\omega^2 - \Omega_+^2}\right] - iE_z\left[\frac{\Omega}{\omega(\omega^2 - \Omega^2)} + \frac{m}{M}\frac{\Omega_+}{\omega(\omega^2 - \Omega_+^2)}\right]\right\} \\
j_z &= -\frac{i\omega}{4\pi}\omega_p^2 \left\{E_z\left[\frac{1}{\omega^2 - \Omega^2} + \frac{m}{M}\frac{1}{\omega^2 - \Omega_+^2}\right] + iE_y\left[\frac{\Omega}{\omega(\omega^2 - \Omega^2)} + \frac{m}{M}\frac{\Omega_+}{\omega(\omega^2 - \Omega_+^2)}\right]\right\}
\end{aligned} \qquad \ldots\ldots(4.4)$$

where

$$\Omega_+ = \frac{e_+ B}{M} = -\frac{m}{M}\Omega.$$

The dielectric coefficient is now a tensor, but may still be defined from Maxwell's equation by

$$\epsilon_{ij} E_j = D_i = E_i + \frac{4\pi j_i}{i\omega},$$

hence (Bernstein 1958)

$$\epsilon = \begin{bmatrix} 1 - \dfrac{\omega_p^2}{\omega^2}\left(1 + \dfrac{m}{M}\right) & 0 & 0 \\ 0 & 1 - \omega_p^2\left(\dfrac{1}{\omega^2 - \Omega^2} + \dfrac{m}{M}\dfrac{1}{\omega^2 - \Omega_+^2}\right) & -i\omega_p^2\left[\dfrac{\Omega}{\omega(\omega^2 - \Omega^2)} + \dfrac{m}{M}\dfrac{\Omega_+}{\omega(\omega^2 - \Omega_+^2)}\right] \\ 0 & i\omega_p^2\left[\dfrac{\Omega}{\omega(\omega^2 - \Omega^2)} + \dfrac{m}{M}\dfrac{\Omega_+}{\omega(\omega^2 - \Omega_+^2)}\right] & 1 - \omega_p^2\left[\dfrac{1}{\omega^2 - \Omega^2} + \dfrac{m}{M}\dfrac{1}{(\omega^2 - \Omega_+^2)}\right] \end{bmatrix}. \qquad \ldots\ldots(4.5)$$

In the dielectric coefficient, fields transverse and parallel to the magnetic fields separate. At high frequencies, $\epsilon \to \mathbf{1}$, and at low frequencies

$$\epsilon \to \begin{bmatrix} -\frac{\omega_p^2}{\omega^2}\left(1+\frac{m}{M}\right) & 0 & 0 \\ 0 & 1+\frac{4\pi n(M+m)}{B^2} & 0 \\ 0 & 0 & 1+\frac{4\pi n(M+m)}{B^2} \end{bmatrix} \qquad \ldots\ldots(4.6)$$

since

$$\omega_p^2\left(\frac{1}{\Omega^2}+\frac{m}{M}\frac{1}{\Omega_+^2}\right) = \frac{4\pi ne^2}{m}\left[\frac{1}{e^2B^2/m^2}+\frac{m}{Me^2B^2/M^2}\right] = \frac{4\pi n(m+M)}{B^2}.$$

The motion of the plasma at these low frequencies is, from (4.2),

$$\frac{e\mathscr{E}}{m}+i\Omega w = 0$$

or
$$E_y+iE_z+B(iv_y-v_z) = 0 = \mathbf{E}+\mathbf{v}\times\mathbf{B}.$$

Thus
$$v_y = -E_z/B, \quad v_z = E_y/B$$

or
$$\mathbf{v} = \mathbf{E}\times\mathbf{B}/B^2. \qquad \ldots\ldots(4.7)$$

These are, however, exactly the motions experienced by a perfectly conducting fluid transmitting Alfvén waves.

The dielectric coefficient may be used to study the propagation of oscillations through the plasma. The problem is complicated by the fact that Maxwell's equations demand $\operatorname{div} \epsilon.\mathbf{E} = 0$ which is not equivalent to $\operatorname{div}\mathbf{E} = 0$; thus the propagation equation has the form for plane waves, where c = velocity of light *in vacuo*

$$\mathbf{k}\times(\mathbf{k}\times\mathbf{E}) = -\frac{\omega^2}{c^2}\epsilon.\mathbf{E}. \qquad \ldots\ldots(4.8)$$

If we write

$$\epsilon = \begin{bmatrix} 1-\alpha & 0 & 0 \\ 0 & 1-\beta & -i\gamma \\ 0 & i\gamma & 1-\beta \end{bmatrix}, \qquad \ldots\ldots(4.9)$$

we obtain a dispersion relation in the form,

$$[n^2-(1-\beta)]\,[n^2\{(1-\alpha)+(\alpha-\beta)\sin^2\theta\}-(1-\alpha)(1-\beta)]+\gamma^2[n^2\sin^2\theta-(1-\alpha)] = 0,$$

where $n = c/v_p = kc/\omega$ and θ is the angle between the direction $\hat{\mathbf{b}}$ of the field and the direction $\hat{\mathbf{k}}$ of propagation, i.e. $\cos\theta = \hat{\mathbf{k}}.\hat{\mathbf{b}}$. This may be solved for the direction of propagation as

$$\sin^2\theta = (1-\alpha)\frac{[\gamma^2-\{n^2-(1-\beta)\}^2]}{n^2\{[n^2-(1-\beta)]\,[\alpha-\beta]+\gamma^2\}}, \qquad \ldots\ldots(4.10)$$

or for the velocity of propagation. The behaviour of this dispersion relation has been analysed by observing that it depends on the three parameters θ, ω/ω_p and ω/Ω, and by examining the critical lines where a resonance $v_p = 0, n = \infty$, or a stop band $v_p = \infty, n = 0$ occurs.

For resonance

$$1-\alpha+(\alpha-\beta)\sin^2\theta = 0$$

or, neglecting the ion motion,

$$\sin^2\theta = \left(\frac{\omega^2}{\omega_p{}^2}-1\right)\left(\frac{\omega^2}{\Omega^2}-1\right), \qquad \ldots\ldots(4.11)$$

while for a stop band

$$(1-\alpha)[(1-\beta)^2-\gamma^2] = 0,$$

i.e. $$\omega^2 = \omega_p{}^2 \quad \text{or} \quad \omega^2 = \omega_p{}^2+\tfrac{1}{2}\Omega^2 \pm \sqrt{[\Omega^2(\omega_p{}^2+\tfrac{1}{4}\Omega^2)]}. \qquad \ldots\ldots(4.12)$$

At low frequencies, $\gamma \to 0$, $\alpha \to \omega_p{}^2/\omega^2$, $\beta \to -4\pi nMc^2/B^2$ and the dispersion relation becomes

$$[n^2-(1-\beta)]\left[n^2\left(\cos^2\theta+\frac{1-\beta}{1-\alpha}\sin^2\theta\right)-(1-\beta)\right] = 0,$$

i.e.

$$v_p = \frac{v_A}{\sqrt{(1+v_A{}^2/c^2)}}$$

where

$$v_A = \sqrt{\left(\frac{B^2}{4\pi\rho}\right)}$$

is the Alfvén speed or

$$v_p = \frac{v_A}{\sqrt{(1+v_A{}^2/c^2)}}\cos\theta.$$

For parallel propagation, $n^2 = (1-\beta) \pm \gamma$, and the waves again propagate with the Alfvén speed, for frequencies far from the resonances, but split into two, one of which slows down as the frequency approaches the gyrofrequency $\omega = \Omega$.

4.2. *The Dielectric Properties of a Hot Magnetized Plasma*

For low-frequency transmission problems the dielectric coefficient (4.5) is an adequate description; however, to discuss the high-frequency, short wavelength behaviour of the plasma, the effect of thermal motions must be included, and we shall now seek the analogue of (3.8) in a magnetized plasma.

To do this we shall linearize the Vlasov equation, but this time keep the term involving the magnetic field and the derivative of the perturbed distribution function (Bernstein 1958). We write

$$\frac{\partial f}{\partial t}+\mathbf{v}.\frac{\partial f}{\partial \mathbf{x}}+\frac{e}{m}\mathbf{v}\times\mathbf{B}.\frac{\partial f}{\partial \mathbf{v}}+\frac{e\mathbf{E}}{m}.\frac{\partial f}{\partial \mathbf{v}} \qquad \ldots\ldots(4.13)$$

and linearize, assuming that a small oscillating field $\mathbf{E}'$ produces a small change f' in the distribution function. Before the field is switched on, the distribution is f_0, which satisfies (4.13) if $\mathbf{E} = 0$; hence the linearized equation becomes

$$\frac{\partial f'}{\partial t}+\mathbf{v}.\frac{\partial f'}{\partial \mathbf{x}}+\frac{e}{m}\mathbf{v}\times\mathbf{B}.\frac{\partial f'}{\partial \mathbf{v}} = -\frac{e\mathbf{E}'}{m}.\frac{\partial f_0}{\partial \mathbf{v}}. \qquad \ldots\ldots(4.14)$$

If we now Fourier analyse $\mathbf{E}'$, $\mathbf{E}'(\mathbf{x}, t) = \Sigma \mathbf{E}(\mathbf{k}, \omega) \exp[i(\mathbf{k}.\mathbf{x} + \omega t)]$, where dropping the superscript should introduce no confusion, the linearized equations for the Fourier components of f' become

$$i(\omega + \mathbf{k}.\mathbf{v}) f'(\mathbf{k}.\omega) + \frac{e}{m}(\mathbf{v} \times \mathbf{B}).\frac{\partial f'}{\partial \mathbf{v}}(\mathbf{k}, \omega) = -\frac{e}{m}\mathbf{E}(\mathbf{k}, \omega).\frac{\partial f_0}{\partial \mathbf{v}} \quad \ldots\ldots(4.15)$$

where f_0 has been assumed stationary and uniform.

To solve (4.15) we may introduce a cylindrical coordinate system to describe $\mathbf{v}$, writing $\mathbf{v} = v_\parallel \mathbf{b} + \mathbf{v}_\perp$, where $\mathbf{v}_\parallel = (\mathbf{v}.\mathbf{B})\mathbf{B}/B^2$ and $\mathbf{v}_\perp = \mathbf{v} - \mathbf{v}_\parallel$, and, further, $\mathbf{v}_\perp = v_\perp(\cos\phi, \sin\phi, 0)$. If we select the origin for ϕ along $\mathbf{k}_\perp$ we obtain

$$i(\omega + k_\parallel v_\parallel + k_\perp v_\perp \cos\phi) f' - \Omega \frac{\partial f'}{\partial \phi} = -\frac{e}{m}\mathbf{E}.\frac{\partial f_0}{\partial \mathbf{v}}. \quad \ldots\ldots(4.16)$$

This has the integrating factor

$$G = \exp\left\{-\frac{i}{\Omega}\int_{\phi_0}^{\phi}[\omega + k_\parallel v_\parallel + k_\perp v_\perp \cos\phi']\, d\phi'\right\}$$
$$= \exp\left\{-\frac{i}{\Omega}[(\omega + k_\parallel v_\parallel)(\phi - \phi_0) + k_\perp v_\perp(\sin\phi - \sin\phi_0)]\right\},$$

where ϕ_0 is arbitrary. If we now insist that f_0 be even in $v_\perp$, we obtain

$$f' = \frac{e}{m\Omega} G(\phi) \int_0^{\phi} d\phi'\, G(\phi') \left\{E_\parallel \frac{\partial f_0}{\partial v_\parallel} + 2\frac{\partial f_0}{\partial v_\perp{}^2}[E_\perp{}' v_\perp \cos(\phi' - \phi_1) + E_\perp{}^2 v_\perp \sin(\phi' - \phi_1)]\right\}.$$

To carry out the integrals here we use the expansion

$$\exp\left\{\frac{ik_\perp v_\perp}{\Omega}\sin\phi\right\} = \sum_n J_n\left(\frac{k_\perp v_\perp}{\Omega}\right) e^{in\phi}$$

whereupon the integral becomes

$$f' = \frac{e}{m} G(\phi) \exp\left\{\frac{i}{\Omega}[(\omega + k_\parallel v_\parallel)\phi_0 + k_\perp v_\perp \sin\phi_0]\right\} \sum_n J_n\left(\frac{k_\perp v_\perp}{\Omega}\right)$$
$$\times \left(E_\parallel \frac{\partial f_0}{\partial v_\parallel} \frac{1}{i(\omega + k_\parallel v_\parallel + n\Omega)} + \frac{\partial f_0}{\partial v_\perp}\left\{\frac{E_\perp{}'}{2i}\left[\frac{1}{\omega + k_\parallel v_\parallel + (n+1)\Omega} + \frac{1}{\omega + k_\parallel v_\parallel + (n-1)\Omega}\right]\right.\right.$$
$$\left.\left. + \frac{E_\perp{}^2}{2}\left[\frac{1}{\omega + k_\parallel v_\parallel + (n+1)\Omega} - \frac{1}{\omega + k_\parallel v_\parallel + (n-1)\Omega}\right]\right\}\right). \quad \ldots\ldots(4.17)$$

To form the current $\mathbf{j}$, we must multiply this by a relevant component of the velocity, then integrate over the phases and the magnitude of the velocity. We obtain

$$j_i = \frac{e^2 n_0}{im} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^{\infty} dv_\parallel\, dv_\perp\, v_\perp A_{ij}{}^n(k_\perp v_\perp) B_{jk}(\omega + n\Omega + k_\parallel v_\parallel) E_k, \quad \ldots\ldots(4.18)$$

where
$$B_{jk}(x) = \begin{bmatrix} \frac{1}{x} & 0 & 0 \\ 0 & \frac{x}{x^2-\Omega^2} & \frac{i\Omega}{x^2-\Omega^2} \\ 0 & \frac{i\Omega}{x^2-\Omega^2} & \frac{x}{x^2-\Omega^2} \end{bmatrix} \qquad \ldots\ldots(4.19)$$

and
$$A_{ij}{}^{n} = -\sum_m J_n\left(\frac{k_\perp v_\perp}{\Omega}\right) J_m\left(\frac{k_\perp v_\perp}{\Omega}\right) a_{jk}{}^{n,m} \qquad \ldots\ldots(4.20)$$

where
$$a_{ij}{}^{n,m} = \begin{bmatrix} \frac{a}{2}[\delta_{n,m}+\delta_{n,m+2}+\delta_{n,m-2}], & \frac{a}{4i}[\delta_{n,m+2}-\delta_{n,m-2}] & \frac{a}{2}[\delta_{n,m+1}+\delta_{n,m-1}] \\ \frac{a}{4i}[\delta_{n,m+2}-\delta_{n,m-2}] & \frac{a}{2}[\delta_{n,m}-\delta_{n,m+2}-\delta_{n,m-2}] & \frac{a}{2i}[\delta_{n,m+1}-\delta_{n,m-1}] \\ \frac{a}{2}[\delta_{n,m+1}+\delta_{n,m-1}] & \frac{a}{2i}[\delta_{n,m+1}-\delta_{n,m-1}] & b\delta_{n,m} \end{bmatrix} \qquad \ldots\ldots(4.21)$$

and
$$a = v_\perp \partial f_0/\partial v_\perp, \quad b = v_\parallel \partial f_0/\partial v_\parallel.$$

The integrals here involve denominators of the form $1/(\omega + k_\parallel v_\parallel + n\Omega) = 1/x$, $1/(x^2-\Omega^2)$, both of which display resonances, and must be handled by the Landau procedure. The three types of singular integrals occurring then must be defined as

$$\int \frac{(\omega-u)g(u)}{(\omega-u)^2-\Omega^2}\,du = -P\int_{-\infty}^{\infty} \frac{ug(u+\omega)\,du}{u^2-\Omega^2} + \frac{i\pi}{2}[g(\omega-\Omega)+g(\omega+\Omega)]$$

$$\int \frac{\Omega g(u)}{(\omega-u)^2-\Omega^2}\,du = P\int_{-\infty}^{\infty} \frac{\Omega g(u+\omega)\,du}{u^2-\Omega^2} - \frac{i\pi}{2}[g(\omega-\Omega)-g(\omega+\Omega)]$$

and
$$\int \frac{g(u)\,du}{\omega-u} = -P\int_{-\infty}^{\infty} \frac{g(u)\,du}{u-\omega} + i\pi g(\omega). \qquad \ldots\ldots(4.22)$$

Of these, the last represents Landau damping, while the first two are associated with energy passed to the electrons (or ions) by a cyclotron resonance. From the dielectric coefficient, dispersion relations may be obtained just as in the case of the cold plasma, and, as in the cold plasma, resonances are found, but now in addition to the resonance at $\omega = \Omega$ further resonances are found at each integral multiple of Ω, for a wave transverse to the plasma. For a wave with a component parallel to the magnetic field these resonances are broadened by Landau damping; however, for waves propagating at right angles to the field no such damping exists.

Various simplifications can be made to the dielectric coefficient. For example, if the transverse wavelength is large compared with the Larmor radii, i.e. if $k_\perp v_\perp/\Omega$ is small, only the zero-order Bessel function contributes, and we obtain ($J_0 \simeq 1$)

$$f' = \frac{e}{m}\frac{G(\phi)}{G(\phi_0)}\left[E_\parallel \frac{\partial f_0}{\partial v_\parallel}\frac{1}{i(\omega+k_\parallel v_\parallel)} + \frac{\partial f_0}{\partial v_\perp}\left\{\frac{E_\perp{}^1}{2i}\left(\frac{1}{\omega+k_\parallel v_\parallel+\Omega}+\frac{1}{\omega+k_\parallel v_\parallel-\Omega}\right)\right.\right.$$
$$\left.\left. + \frac{E_\perp{}^2}{2}\left(\frac{1}{\omega+k_\parallel v_\parallel+\Omega}-\frac{1}{\omega+k_\parallel v_\parallel-\Omega}\right)\right\}\right] \qquad \ldots\ldots(4.23)$$

and
$$j_1 = \frac{e^2}{m}\int v_1 E_1 \frac{\partial f_0}{\partial v_1}\frac{1}{i(\omega + k_1 v_1)}\,d^3v$$

$$j_\perp{}^1 = \frac{e^2}{m}\int v_\perp \frac{\partial f_0}{\partial v_\perp}\left[\frac{(\omega + k_1 v_1)E_\perp{}^1}{(\omega + k_1 v_1)^2 - \Omega^2} + \frac{i\Omega E_\perp{}^2}{(\omega + k_1 v_1)^2 - \Omega^2}\right]$$

$$j_\perp{}^2 = \frac{e^2}{m}\int v_\perp \frac{\partial f_0}{\partial v_\perp}\left[\frac{(\omega + k_1 v_1)E_\perp{}^2}{(\omega + k_1 v_1)^2 - \Omega^2} - \frac{i\Omega E_\perp{}^1}{(\omega + k_1 v_1)^2 - \Omega^2}\right].$$

In this approximation the dielectric coefficient in the magnetic field is modified by temperature effects exactly as was the dielectric coefficient in absence of a field.

4.3. *Plasma in a Non-uniform Field; Approximation Methods*

The properties of the dielectric coefficient in a magnetic field are extremely complex, and only a few of the grosser effects have so far had much application. It is clear from this that if we are to study the properties of a plasma in a non-uniform, non-steady magnetic field some approximation procedure is required. For very high frequency phenomena, the relevant processes are usually local, and these may be discussed by using geometrically simplified models; for example, in electromagnetic transmission problems, the dielectric coefficient may be considered as a slowly varying function of position. On the other hand, low frequency processes which occur on a large scale and involve the geometrical structure in an essential way can be handled by using approximate methods to describe the particle motions.

As a preliminary we must discuss the approximate motion of particles in magnetic fields which vary slowly in space and time, that is in which the field changes little during a single rotation of the particle in its Larmor orbit, i.e.

$$\frac{1}{\Omega}\frac{1}{B}\frac{\partial B}{\partial t} \ll 1 \quad \text{and} \quad r_L \frac{1}{B}\frac{\partial B}{\partial x} \ll 1$$

where $\Omega = eB/m$ is the cyclotron frequency and $r_L = v_\perp/\Omega$ the Larmor radius. In a stationary uniform magnetic field a particle moves freely along a line of force, while rotating about the line of force with an angular frequency Ω at a radius r_L. In slowly varying fields, the motion is only slightly altered; particles still move in almost circular orbits about a guiding centre, which drifts almost freely along field lines. Because of the variation of the fields, however, the motion is no longer strictly periodic, and in particular a steady drift velocity is produced across the magnetic field lines. We have already observed one component of this drift, that induced by an electric field normal to the magnetic field, $\mathbf{V}_E = (\mathbf{E}\times\mathbf{B})/B^2$, but any non-electric acceleration will produce a similar drift. If the equations of motion are written as

$$\dot{\mathbf{V}} = \mathbf{F} + \frac{e}{m}[\mathbf{E} + \mathbf{v}\times\mathbf{B}], \qquad \ldots\ldots(4.24)$$

it is possible to introduce a drift velocity $\mathbf{V}_D$

$$\mathbf{V}_D = \left(\frac{m}{e}\mathbf{F} + \mathbf{E}\right)\times\mathbf{B}/B^2, \qquad \ldots\ldots(4.25)$$

so that
$$\frac{m}{e}\mathbf{F}+\mathbf{E}+\mathbf{V}_D\times\mathbf{B}=\frac{m}{e}\mathbf{F}_\parallel+\mathbf{E}_\parallel, \qquad \ldots\ldots(4.26)$$

where $\mathbf{A}_\parallel=(\mathbf{A}.\hat{\mathbf{b}})\hat{\mathbf{b}}$, $\mathbf{A}_\perp=\mathbf{A}-\mathbf{A}_\parallel$, $\mathbf{B}=B\hat{\mathbf{b}}$, and a relative velocity $\mathbf{c}$ such that $\mathbf{c}=\mathbf{v}-\mathbf{V}_D$ and

$$\dot{\mathbf{c}}=\frac{e}{m}\mathbf{c}\times\mathbf{B}+\mathbf{F}_\parallel+\frac{e}{m}\mathbf{E}_\parallel+\dot{\mathbf{V}}_D. \qquad \ldots\ldots(4.27)$$

We may now make use of the approximation of slowly varying fields, by introducing an expansion parameter ϵ. When (4.27) is made dimensionless by the introduction of some measure c_0 of the relative velocity c, and the gyrofrequency eB/m, this requires that

(i) $\dfrac{F_\parallel+eE_\parallel/m}{\Omega c_0}\simeq O(\epsilon)$, (ii) $\dfrac{1}{\Omega}\dfrac{\partial}{\partial t}$ and $r_L\dfrac{\partial}{\partial x}$ on $\dfrac{m}{e}\mathbf{F}$, $\mathbf{E}$ and $\mathbf{B}$ be of order ϵ.

Thus if, in addition to the condition that the fields be slowly varying, we insist that the component of acceleration along the magnetic field be small, we may reduce (4.27) to the form

$$\dot{\mathbf{c}}=\Omega\mathbf{c}\times\hat{\mathbf{b}}+\epsilon\mathbf{f}(\mathbf{r},t), \qquad \ldots\ldots(4.28)$$

where Ω and $\hat{\mathbf{b}}$ themselves may vary to order ϵ. Our entire approximation procedure will be based on a solution $\mathbf{c}$ expanded in powers of ϵ,

$$\mathbf{c}=\mathbf{c}_0+\epsilon\mathbf{c}_1+\epsilon^2\mathbf{c}_2+\ldots.$$

In zero order $\mathbf{c}$ is periodic with period $2\pi/\Omega$; hence the motion relative to the guiding centre R is given by

$$\ddot{\mathbf{r}}+\Omega^2(\mathbf{r},t)\,\mathbf{r}=\epsilon\mathbf{g}(\mathbf{r},t), \qquad \ldots\ldots(4.29)$$

where $\Omega=\Omega_0+\epsilon\Omega_1+\ldots$. For any function r which may be expanded as $r_0(t)+\epsilon r_1(t)+\ldots$ where $r_0(t)$ is periodic there exists an important adiabatic invariant $\alpha=\oint\dot{r}\,dr$, where the integral extends around the period of r_0. This has the important property that $\dot{\alpha}(r_0)=O(\epsilon^2)$, for

$$\dot{\alpha}(r_0+\epsilon r_1)=\frac{d}{dt}\oint(\dot{r}_0+\epsilon\dot{r}_1)^2\,dt=\frac{d}{dt}\oint\dot{r}_0^2\,dt+\epsilon\oint(\ddot{r}_0\dot{r}_1+\dot{r}_0\ddot{r}_1)\,dt+O(\epsilon^2)=\dot{\alpha}(r_0)+O(\epsilon^2),$$

for on integrating the third term by parts, noting that the periodicity of r_0 implies the vanishing of $\dot{r}_0$ at the limits,

$$\epsilon\oint(\ddot{r}_0\dot{r}_1+\dot{r}_0\ddot{r}_1)\,dt=\epsilon\dot{r}_0\dot{r}_1\Big|+\oint[-\dot{r}_0\ddot{r}_1+\dot{r}_0\ddot{r}_1]\,dt=0,$$

the integrated term vanishing from the periodicity of r_0.

However, $\alpha(r_0)=\oint\dot{r}_0dr_0=\Omega\oint(dr/d\theta)^2\,d\theta=\frac{1}{2}\Omega a^2$ where a is the amplitude of r_0. Here we do not insist that $\Omega(t)$ be constant, but $\dot{\Omega}(t)\simeq O(\epsilon)$. For a charged particle in a magnetic field

$$\alpha=\tfrac{1}{2}\Omega r_L^2=\tfrac{1}{2}\frac{mV_\perp^2}{eB(t)}, \qquad \ldots\ldots(4.30)$$

where B is evaluated on the guiding centre, is an adiabatic invariant. We may identify $\frac{1}{2}[mV_\perp{}^2/B(t)]$ with μ, the effective dipole moment produced by the charged particle travelling in small circles in the magnetic field, a quantity which has been exploited by Alfvén.

To proceed with a formal solution to the equations of motion, it is convenient to represent the velocity by its projections on a triad of unit vectors $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3$ which incorporate some information about the geometry of the magnetic field, thus

$$\hat{\mathbf{e}}_1 = \hat{\mathbf{b}}, \quad \frac{\hat{\mathbf{e}}_2}{R} = \frac{\partial \hat{\mathbf{b}}}{\partial x_1}, \quad \hat{\mathbf{e}}_1 \times \hat{\mathbf{e}}_2 = \hat{\mathbf{e}}_3, \quad \hat{\mathbf{e}}_2 \times \hat{\mathbf{e}}_3 = \hat{\mathbf{e}}_1, \quad \hat{\mathbf{e}}_3 \times \hat{\mathbf{e}}_1 = \hat{\mathbf{e}}_2. \qquad \ldots\ldots(4.31)$$

From
$$0 = \operatorname{div} \mathbf{B} = \operatorname{div}(B\hat{\mathbf{e}}_1) = \hat{\mathbf{e}}_1 . \nabla B + \mathbf{B} \operatorname{div} \hat{\mathbf{e}}_1,$$
$$\operatorname{div} \hat{\mathbf{e}}_1 = -\hat{\mathbf{e}}_1 . \nabla \log B. \qquad \ldots\ldots(4.32)$$

In terms of these vectors

$$\dot{\mathbf{c}} = \hat{\mathbf{e}}_1 \dot{c}_1 + \hat{\mathbf{e}}_2 \dot{c}_2 + \hat{\mathbf{e}}_3 \dot{c}_3 + \dot{\hat{\mathbf{e}}}_1 c_1 + \dot{\hat{\mathbf{e}}}_2 c_2 + \dot{\hat{\mathbf{e}}}_3 c_3,$$

where
$$\dot{\hat{\mathbf{e}}} = \frac{\partial \hat{\mathbf{e}}}{\partial t} + (\mathbf{V}_D + \mathbf{c}) . \nabla \hat{\mathbf{e}}. \qquad \ldots\ldots(4.33)$$

We are particularly interested in the first-order aperiodic contributions, hence we require only the aperiodic contributions from the first-order terms. These are

$$\left(\frac{D}{Dt}\hat{\mathbf{e}}_1\right) c_1 + \hat{\mathbf{e}}_2 \frac{c_1{}^2}{2R} + \left(\frac{\partial \hat{\mathbf{e}}_2}{\partial r_2} + \frac{\partial \hat{\mathbf{e}}_3}{\partial r_3}\right) \tfrac{1}{2} c_{\perp 0}{}^2$$

where
$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{V}_D . \nabla.$$

Using $\hat{\mathbf{e}}_2 = \hat{\mathbf{e}}_3 \times \hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_3 = \hat{\mathbf{e}}_1 \times \hat{\mathbf{e}}_2$ the last term may be reduced to

$$\hat{\mathbf{e}}_1 \tfrac{1}{2} c_\perp{}^2 \operatorname{div} \hat{\mathbf{e}}_1 = -\hat{\mathbf{e}}_1 \frac{1}{2} \frac{c_\perp{}^2}{B} \frac{\partial B}{\partial x_1}. \qquad \ldots\ldots(4.34)$$

The right-hand side of (4.27) may be treated in a similar fashion. $\dot{\mathbf{V}}_D$ reduces to $(D/Dt)\mathbf{V}_D$, hence on integration becomes $\mathbf{V}_D$, our approximation; hence if we subtract this term and define $\mathbf{c}$ with respect to the instantaneous guiding centre motion, it contributes no further. The term in Ω must be broken up into $\Omega(c_1 t, 0, 0) + (\partial\Omega/\partial r_\perp{}^0)\, \mathbf{r}_\perp{}^0$ where $\mathbf{r}_\perp{}^0 = \int c_\perp{}^0 \, dt$ is bounded since $c_\perp{}^0$ is periodic. The three equations correct to first order become

$$\dot{c}_1 = \epsilon \left[\left(F + \frac{e}{m} E_1\right) - \frac{1}{2} \frac{c_\perp{}^2}{B} \frac{\partial B}{\partial x_1}\right]$$

$$\dot{c}_2 = -\Omega c_3 - \epsilon \left[\frac{\partial \Omega}{\partial r_2} r_2{}^0 c_3{}^0 - \hat{e}_2 c_1 \frac{D\hat{e}_1}{Dt} + \frac{c_1{}^2}{R}\right]$$

$$\dot{c}_3 = \Omega c_2 + \epsilon \left[\frac{\partial \Omega}{\partial r_3} r_3{}^0 c_2{}^0 - \hat{e}_3 c_1 \frac{D\hat{e}_1}{Dt}\right]. \qquad \ldots\ldots(4.35)$$

To order ϵ then c_1 is not constant, but experiences in addition to the acceleration $F_1 + eE_1/m$ a further $-\frac{1}{2}(c_\perp{}^2/B)/(\partial B/\partial x_1)$, which represents the interaction of a field gradient and the effective dipole moment $\mu = \frac{1}{2}(mv_\perp{}^2/B)$ of the particle. From

the terms of order ϵ in $\dot{c}_2$ and $\dot{c}_3$ we obtain three further drift velocities, one due to the motion of the flux lines

$$\frac{1}{\Omega}\frac{D\hat{\mathbf{e}}_1}{Dt}\times\hat{\mathbf{e}}_1$$

one due to the curvature

$$\frac{c_\parallel{}^2}{\Omega}\frac{\partial\hat{\mathbf{e}}_1}{\partial x_\parallel}\times\hat{\mathbf{e}}_1=\frac{c_\parallel{}^2}{\Omega R}\hat{\mathbf{e}}_3$$

and one due to the variation in field strength

$$\frac{c_\perp{}^2}{2\Omega}\hat{\mathbf{e}}_1\times\nabla\log B.$$

If the forces are stationary, there is an energy integral arising from $\dot{c}_\parallel$,

$$c_\parallel{}^2+c_\perp{}^2-\frac{1}{2}\int\left(F_\parallel+\frac{e}{m}E_\parallel\right)dx_\parallel=\beta=\text{constant}. \qquad \ldots\ldots(4.36)$$

From these approximate solutions to the equations of motion a number of inferences can be made. If there are no non-electric forces, then particles experience a charge-independent drift $\mathbf{V}_E$, and to zero order in ϵ, the system behaves as a perfect conductor

$$\mathbf{E}+\mathbf{V}_E\times\mathbf{B}=0. \qquad \ldots\ldots(4.37)$$

The presence of a non-electric acceleration introduces a current $\mathbf{j}=\Sigma(m\mathbf{F}\times\mathbf{B}/B^2)$; but if the acceleration is independent of the mass, no further drift; thus a plasma acted upon by only gravitational and electric forces behaves like a perfect conductor to order ϵ. In order ϵ, the first-order drifts introduce further currents, but also further mass velocities, and the perfect conduction hypothesis becomes invalid.

The constancy of the magnetic moment is a most important result. It is the basis of the open-ended magnetic mirror approach to plasma containment (Post 1956). A magnetic mirror differs from the magnetohydrodynamic devices in having the field go out of the plasma. As it leaves the plasma in either direction its strength increases. Since $c_\perp{}^2/B=\text{const}=\alpha$, $c_\parallel$ will reach zero at a point where the field strength has increased over its minimum value B_0 to a value $B_{\max}$ such that

$$\frac{c_{\perp 0}{}^2}{B_0}=\frac{c_{\parallel 0}{}^2+c_{\perp 0}{}^2}{B_{\max}}, \quad \text{or} \quad \frac{c_{\perp 0}{}^2}{c_{\parallel 0}{}^2+c_{\perp 0}{}^2}=\frac{B_0}{B_{\max}};$$

thus, any particle whose velocity at B_0 makes an angle $\theta>\theta_0$ with the field line where $\sin^2\theta_0=B_0/B_{\max}$ will be unable to pass $B_{\max}$, and will be trapped. The identical mechanism is involved in Fermi's theory (1949) of the acceleration of cosmic rays; particles make mirror reflections with magnetized gas clouds in interstellar space and acquire energy by the excess of head-on over head-to-tail collisions.

4.4. *Some Exact Solutions for the Steady State*

Before proceeding to an elaborate discussion of the approximate solutions to a wide class of problems, we will consider a special class for which the Vlasov equation is exactly soluble. These depend upon finding exact constants of motion and are confined to symmetric stationary systems. Clearly, an arbitrary function

of the constants of motion satisfies the Vlasov equation, since it says no more than $Df/Dt|_{\text{traj}} = 0$ where the total derivative follows particle orbits and for a constant of the motion α, $D\alpha/Dt|_{\text{traj}} = 0$. To find exact constants of the motion, it is convenient to consider the equations of motion in Hamiltonian form,

$$H = \frac{1}{2m}(p_i - eA_i)^2 + e\phi, \qquad \ldots\ldots(4.38)$$

where p_i and A_i are the components of the canonical momentum and the vector potential, while ϕ is the scalar potential. If H is independent of some coordinate q_j, the corresponding canonical momentum is constant, since

$$\dot{p}_j = -\frac{\partial H}{\partial q_j} = 0.$$

The momenta are determined from the velocities by

$$\dot{q}_i = (p_i - eA_i)/m.$$

H itself is a constant of the motion.

As a simple example, consider a cylindrically symmetric system (Sparrow 1960, Harris 1957) in which an axial current flows so that $A_z \neq 0$ and consider a distribution function

$$\begin{aligned} f &= C\exp\{-\beta[H - wp_z]\} \\ &= C\exp\left\{-\beta\left[\tfrac{1}{2}v_\perp^2 + \tfrac{1}{2}v_z^2 + \frac{e\phi}{m} - w\left(v_z + \frac{e}{m}A_z\right)\right]\right\}. \qquad \ldots\ldots(4.39) \end{aligned}$$

The constant C is determined by setting the density $n = n_0$ at $A = 0$, $\phi = 0$, whereupon

$$C = n_0\left(\frac{\beta}{2\pi}\right)^{3/2}\exp(-\tfrac{1}{2}\beta w^2) \quad \text{and} \quad n = n_0\exp[-\beta(e\phi - weA_z)/m].$$

The drift velocity

$$\begin{aligned} n\langle v_z\rangle &= n\left(\frac{\beta}{2\pi}\right)^{3/2}\int\exp\{-\beta[\tfrac{1}{2}v_\perp^2 + \tfrac{1}{2}(v_z - w)^2]\}\, v_z\, d^3v \\ &= nw. \end{aligned}$$

To complete the problem the fields must be determined, and we must introduce two distribution functions for the ions and electrons, thus

$$n_- = n_0\exp\left[\frac{\beta_-}{m_-}e(\phi - wA_z)\right], \quad n_+ = n_0\exp\left(-\frac{\beta_+}{m_+}e\phi\right)$$

and insisting $n_+ = n_-$ determines $\phi(A_z)$ as

$$e\phi = \frac{\beta_- ewA_z}{\beta_- + (m_-/m_+)\beta_+},$$

hence

$$n_- = \exp\left\{\frac{-m_-m_+\beta_-\beta_+ ewA_z}{m_+\beta_- + m_-\beta_+}\right\} \equiv \exp(\alpha A_z).$$

Finally, Maxwell's equation determining the vector potential becomes

$$\frac{1}{r}\frac{\partial}{\partial r} r \frac{\partial A_z}{\partial r} = -4\pi j_z = 4\pi n_0 e w \exp(-\alpha A z)$$

and introducing $y = \exp(-\alpha A_z)$, $x^2 = 4\pi n_0 e w \alpha r^2$,

$$\frac{1}{x}\frac{d}{dx} x \frac{dy}{dx} + y = 0$$

and if $A_z(0) = 0$, $y(0) = 1$, $y'(0) = 0$,

$$y = 1/(1 + \tfrac{1}{8}x^2)^2.$$

Finally if $\beta_- = m_-/kT_-$, $\beta_+ = m_+/kT_+$,

$$n = n_0 \Big/ \left\{1 + \frac{\pi n_0 e^2}{2m} \frac{mw^2}{k(T_+ + T_-)} r^2\right\}^2, \qquad \ldots\ldots(4.40)$$

a solution given by Bennett (1934) and independently by Thonemann and Cowhig (1951). There are other exact solutions which correspond to a rigid rotation of the cylinder while a number of plane solutions are known.

4.5. *Approximate Procedure for solving the Vlasov Equation*

We are now ready to develop an approximate method for solving Vlasov's equation in slowly varying fields (Vlasov 1938). In the procedure sketched here the properties of the motion do not appear to play an important part; however, when attempts are made to solve the approximate equations the importance of the adiabatic invariants of motion becomes obvious. To generate the approximations we first write Vlasov's equation in a normalized form exactly as Boltzmann's equation was written in (§ 2)

i.e.
$$(\mathbf{E}' + \mathbf{v}' \times \mathbf{B}') \cdot \frac{\partial f}{\partial \mathbf{v}'} + \frac{1}{\Omega_0 T}\left(\frac{\partial f}{\partial t'} + \mathbf{v}' \cdot \frac{\partial f}{\partial \mathbf{x}'}\right) = 0, \qquad \ldots\ldots(4.41)$$

where in addition to the characteristic length L, time T, and velocity $v_0 = L/T$, a characteristic magnetic field B_0, $\Omega_0 = eB_0/m$, is introduced, and

$$\mathbf{B}' = \mathbf{B}/B_0, \quad \mathbf{E}' = \mathbf{E}v_0/B_0.$$

Our first expansion is based on the assumption that $1/\Omega_0 T = \lambda$ is small. To carry out the procedure, however, we must make the further assumption that $E_\parallel$ is small of order λ and write:

$$(\mathbf{E}_\perp' + \mathbf{v}' \times \mathbf{B}') \cdot \frac{\partial f}{\partial \mathbf{v}'} + \lambda\left(\frac{\partial f}{\partial t'} + \mathbf{v}' \cdot \frac{\partial f}{\partial \mathbf{x}'} + \frac{eE_\parallel}{\lambda}\frac{\partial f}{\partial v_\parallel}\right).$$

We shall seek a solution by expanding f in powers of λ, $f = f_0 + \lambda f^1 + \ldots \lambda^n f^n$, and obtain a set of equations

$$(\mathbf{E}_\perp' + \mathbf{v}' \times \mathbf{B}') \cdot \frac{\partial f^n}{\partial \mathbf{v}'} = -\left(\frac{\partial}{\partial t'} + \mathbf{v}' \cdot \frac{\partial}{\partial \mathbf{x}'} + \frac{eE_\parallel}{\lambda}\frac{\partial}{\partial v_\parallel}\right) f^{n-1}. \qquad \ldots\ldots(4.42)$$

Since we shall need the inverse of the operator on the left here, it is profitable to make a series of changes on the independent variable to simplify this. Using (4.41)

in un-normalized form, we introduce first two new quantities $\mathbf{c}$, an independent variable, and $\mathbf{V}(\mathbf{x}, t)$ such that

$$\mathbf{v} = \mathbf{V} + \mathbf{c} \quad \text{and} \quad \mathbf{E} + \mathbf{V} \times \mathbf{B} = 0, \qquad \ldots\ldots(4.43)$$

thereby reducing the left-hand side of (4.41) to $(e/m)\,\mathbf{c} \times \mathbf{B} . (\partial f/\partial \mathbf{c})$, at the expense of complicating the right-hand side. For $\partial f/\partial \mathbf{x}, \partial f/\partial t$ are taken at fixed $\mathbf{v}$, but since

$$\left.\frac{\partial f}{\partial t}\right|_{\mathbf{v}} = \left.\frac{\partial f}{\partial t}\right|_{\mathbf{c}} + \frac{\partial f}{\partial \mathbf{c}} \cdot \left.\frac{\partial \mathbf{c}}{\partial t}\right|_{\mathbf{v}} = \left.\frac{\partial f}{\partial t}\right|_{\mathbf{c}} - \frac{\partial f}{\partial \mathbf{c}} \cdot \frac{\partial \mathbf{V}}{\partial t}$$

we obtain, on the right

$$-\left[\frac{\partial}{\partial t} + (\mathbf{V}+\mathbf{c}) \cdot \frac{\partial}{\partial \mathbf{x}} + eE_{\parallel} \frac{\partial}{\partial c_{\parallel}} - \left\{\left(\frac{\partial}{\partial t} + (\mathbf{V}+\mathbf{c}) \cdot \frac{\partial}{\partial \mathbf{x}}\right)\mathbf{V}\right\} \cdot \frac{\partial}{\partial \mathbf{c}}\right] f .$$

Next, instead of fixed components $\mathbf{c}$, we introduce the projections along and across the magnetic field lines by $c_{\parallel} = \mathbf{c} . \mathbf{b}$, $\mathbf{c}_{\perp} = \mathbf{c} - (\mathbf{c} . \mathbf{b})\mathbf{b}$ and $[\partial/\partial t + (\mathbf{V}+\mathbf{c}) . \partial/\partial \mathbf{x}]|_{\mathbf{c}} f$ becomes

$$\left(\frac{\partial}{\partial t} + (\mathbf{V}+\mathbf{c}) \cdot \frac{\partial}{\partial \mathbf{x}}\right)\Big|_{c_{\parallel}\,\mathbf{c}_{\perp}} f - \mathbf{c} . \left[\left(\frac{\partial}{\partial t} + (\mathbf{V}+\mathbf{c}) \cdot \frac{\partial}{\partial \mathbf{x}}\right)\mathbf{b}\right] \frac{\partial f}{\partial c_{\parallel}} + c_{\parallel}\left[\left(\frac{\partial}{\partial t} + (\mathbf{V}+\mathbf{c}) \cdot \frac{\partial}{\partial \mathbf{x}}\right)\mathbf{b}\right] \cdot \frac{\partial f}{\partial \mathbf{c}_{\perp}} .$$

Finally, we introduce a polar representation of $\mathbf{c}_{\perp}$; $\mathbf{c}_{\perp} = c_{\perp}(\cos\phi, \sin\phi)$. This reduces the left-hand side to $-(eB/m)(\partial f/\partial \phi)$, and the successive approximate equations become

$$\frac{\partial f^n}{\partial \phi} = \frac{1}{\Omega(\mathbf{x}, t)} \mathscr{D} f^{n-1}, \qquad \ldots\ldots(4.44)$$

where, if

$$D' = \frac{\partial}{\partial t} + (\mathbf{V}+\mathbf{c}) . \frac{\partial}{\partial \mathbf{x}},$$

$$\mathscr{D} f = D' f - D' V_i \frac{\partial f}{\partial c_i} - c_i (D' \mathbf{b})_i \frac{\partial f}{\partial c_{\parallel}} + c_{\parallel} (D' \mathbf{b})_i \frac{\partial f}{\partial c_i} . \qquad \ldots\ldots(4.45)$$

We must now seek solutions of (4.41) which are periodic in ϕ. The zero-order solution is simply

$$f_0 = f_0(c_{\perp}^2, c_{\parallel}\,;\ \mathbf{x}, t) \qquad \ldots\ldots(4.46)$$

and the first-order solution

$$f_1 = h(c_{\perp}^2, c_{\parallel}\,;\ \mathbf{x}, t) + \frac{1}{\Omega} \int \mathscr{D} f_0 \, d\phi . \qquad \ldots\ldots(4.47)$$

Since f_1 must be periodic in ϕ, a constraint is placed upon f_0, namely

$$[\mathscr{D} f_0] = \frac{1}{2\pi} \int_0^{2\pi} \mathscr{D} f_0 \, d\phi = 0. \qquad \ldots\ldots(4.48)$$

This constraint may be written

$$\frac{\partial f_0}{\partial t} + \mathbf{V} . \frac{\partial}{\partial \mathbf{x}} f_0 + c_{\parallel} \frac{\partial f_0}{\partial x_{\parallel}} - (\operatorname{div} V)_{\perp} c_{\perp}^2 \frac{\partial f_0}{\partial c_{\perp}^2} - (\operatorname{div} V)_{\parallel} c_{\parallel} \frac{\partial f}{\partial c_{\parallel}} + \frac{e}{m} E_{\parallel} \frac{\partial f_0}{\partial c_{\parallel}}$$

$$- c_{\parallel} c_{\perp}^2 \frac{1}{B} \frac{\partial B}{\partial x_{\parallel}} \left(\frac{\partial f_0}{\partial c_{\parallel}^2} - \frac{\partial f_0}{\partial c_{\perp}^2}\right) = 0, \qquad \ldots\ldots(4.49)$$

and the inhomogeneous part of f_1 written

$$f_1^{i} = \frac{1}{\Omega}\left\{c_y\left[\frac{\partial f}{\partial x}+\frac{\partial f}{\partial c_\parallel}D^* b_x - 2\frac{\partial f}{\partial c_\perp{}^2}(D^* V_x + c_\parallel D^* b_x)\right]\right.$$
$$- c_x\left[\frac{\partial f}{\partial y}+\frac{\partial f}{\partial c_\parallel}D^* b_y - 2\frac{\partial f}{\partial c_\perp{}^2}(D^* V_y + c_\parallel D^* b_y)\right]$$
$$+\tfrac{1}{2}c_x c_y\left[\frac{\partial f}{\partial c_\parallel}\left(\frac{\partial b_x}{\partial x}-\frac{\partial b_y}{\partial y}\right)-2\frac{\partial f}{\partial c_\perp{}^2}\left(\frac{\partial V_x}{\partial x}-\frac{\partial V_y}{\partial y}-c_\parallel\frac{\partial b_x}{\partial x}+c_\parallel\frac{\partial b_y}{\partial y}\right)\right]$$
$$\left.+\tfrac{1}{4}(c_y{}^2-c_x{}^2)\left[\frac{\partial f}{\partial c_\parallel}\left(\frac{\partial b_x}{\partial y}+\frac{\partial b_y}{\partial x}\right)-2\frac{\partial f}{\partial c_\perp{}^2}\left(\frac{\partial V_x}{\partial y}+\frac{\partial V_y}{\partial x}\right)\right]\right\} \qquad \ldots\ldots(4.50)$$

where
$$D^* = \frac{\partial}{\partial t}+\mathbf{V}.\nabla + c_\parallel\frac{\partial}{\partial x_\parallel}.$$

At each stage in the solution procedure, a similar constraint must be applied, which serves to determine the homogeneous solution; and a similar inhomogeneous solution is found.

4.6. *The Magnetohydrodynamic Equations*

We are now in a position to form the magnetohydrodynamic equations. These may be obtained by forming the moments of (4.49) with m, and $m\mathbf{v}$, then adding the contributions for ions and electrons. When this is done we obtain two equations of continuity

$$\frac{\partial \rho}{\partial t}+\operatorname{div}\rho\mathbf{V}=0,$$

where
$$\mathbf{V}=\left(\frac{m}{e}\mathbf{F}+\mathbf{E}\right)\times\frac{\mathbf{B}}{B^2}.$$

If there are no non-electric forces, or if these are weak, so that $\mathbf{F}/\Omega$ is a first-order quantity, the zero-order velocities of ions and electrons are equal, and in zero order

$$\mathbf{E}+\mathbf{V}\times\mathbf{B}=0. \qquad \ldots\ldots(4.51)$$

That is, if not only the gradients are small, but the acceleration experienced by particles is small, then the plasma behaves as a perfect conductor.

The only component of the acceleration coming from (4.49) is

$$\rho\frac{DV_\parallel}{Dt}=-\frac{\partial p_\parallel}{\partial x_\parallel}-\alpha(p_\perp-p_\parallel)+qE_\parallel \qquad \ldots\ldots(4.52)$$

where $\alpha=(1/B)(\partial B/\partial x_\parallel)$ for the electrons and the ions. If the ion and electron pressures are everywhere equal, quasi-neutrality requires $E_\parallel=0$.

The transverse component of acceleration must be obtained using the inhomogeneous f_1.

For each gas we obtain

$$\rho v_x{}^{(1)}=-\frac{1}{\Omega}\left[\frac{\partial}{\partial y}\hat{p}_\perp+\rho\frac{DV_y}{Dt}+\hat{p}_\parallel\frac{\partial b^y}{\partial x_\parallel}\right]$$
$$\rho v_y{}^{(1)}=\frac{1}{\Omega}\left[\frac{\partial}{\partial x}\hat{p}_\perp+\rho\frac{DV_x}{Dt}+\hat{p}_\parallel\frac{\partial b^x}{\partial x_\parallel}\right] \qquad \ldots\ldots(4.53)$$

which yields when the contributions are added

$$\rho \frac{D\mathbf{V}}{Dt} = -\nabla p_\perp + \mathbf{j} \times \mathbf{B} - \frac{\mathbf{n}}{R} p_\parallel \qquad \ldots\ldots(4.54)$$

where R is the radius of curvature and $\mathbf{n}$ the principal normal to the field line.

These relations are identical with the magnetohydrodynamic relations provided we introduce the pressure tensor $\mathbf{p}$ which has the form

$$p_\perp \mathbf{1} + (p_\parallel - p_\perp)\hat{\mathbf{b}}.\hat{\mathbf{b}}, \qquad \ldots\ldots(4.55)$$

i.e.
$$p_{ij} = p_\perp(\delta_{ij} - \hat{b}_i \hat{b}_j) + p_\parallel \hat{b}_i \hat{b}_j$$

where $\hat{\mathbf{b}}$ is the unit vector along the magnetic field and $p_\perp$ and $p_\parallel$ are two distinct scalars.

4.7. *The Equations of State*

The equations (Chew, Goldberger and Low 1956) determining $\mathbf{p}$ may also be formed from f_0, by forming the moments with $\frac{1}{2}mc_\perp^2, mc_\parallel^2$. In general these involve the heat conduction along the field lines $q = \langle \frac{1}{2}mc_\perp^2 c_\parallel \rangle$, and the velocity $V_\parallel$ parallel to the field. We cannot calculate these quantities without producing actual solutions to the Vlasov equation ; however, as we shall see, there are many situations in which the zero-order distribution function is even in $c_\parallel$, and motion occurs solely across the lines of force. In this case we obtain the equations of state

$$\frac{Dp_\perp}{Dt} + 2p_\perp(\nabla . \mathbf{V}) = 0, \quad \frac{Dp_\parallel}{Dt} + p_\parallel(\nabla . \mathbf{V}) = 0. \qquad \ldots\ldots(4.56)$$

Since $\nabla . \mathbf{V} = -(1/\rho)(D\rho/Dt)$, these may be written in the double adiabatic form

$$\frac{D}{Dt}\left(\frac{p_\perp}{\rho^2}\right) = 0, \quad \frac{D}{Dt}\left(\frac{p_\parallel}{\rho}\right) = 0. \qquad \ldots\ldots(4.57)$$

4.8. *Transverse Transport Coefficients*

With our present solution it is also not difficult to get some of the transverse transport coefficients. As we shall see, the solution depends on non-local effects along the lines of force, hence we can scarcely expect to define local quantities except for the transverse components of heat flow, and the pressure tensor. Electrical conductivity depends on re-defining the drift velocities and so induces complications. The off-diagonal components of the stress tensor, however, are easily found by forming the moments

$nm\langle c_\parallel c_\perp \cos\phi \rangle = p_{zx}$, $nm\langle c_\parallel c_\perp \sin\phi \rangle = p_{zy}$ and $nm\langle c_\perp^2 \cos\phi \sin\phi \rangle = p_{xy} = p_{yx}$

using f_1. We discover

$$p_{zx} = -\frac{p_\perp}{\Omega}\left(\frac{\partial V_y}{\partial z} + \frac{\partial V_z}{\partial y}\right),$$

$$p_{zy} = \frac{p_\perp}{\Omega}\left(\frac{\partial V_x}{\partial z} + \frac{\partial V_z}{\partial x}\right),$$

$$p_{xy} = \frac{p_\perp}{2\Omega}\left(\frac{\partial V_x}{\partial x} - \frac{\partial V_y}{\partial y}\right). \qquad \ldots\ldots(4.58)$$

Since $\Omega_+ \ll \Omega_-$ only the ions contribute effectively to this. Slightly less straightforward is the calculation of the heat transfer. There are now several terms which we can calculate, such as $\mathbf{q}_\perp{}^\perp = \langle n\tfrac{1}{2}mc_\perp{}^2\mathbf{c}_\perp\rangle$, $\mathbf{q}_\perp{}^\parallel = \langle n\tfrac{1}{2}mc_\parallel{}^2\mathbf{c}_\perp\rangle$. For each gas these appear as

$$\mathbf{q}_\perp{}^\perp = \frac{1}{\Omega}\left\{\mathbf{b}\times\nabla nm\tfrac{1}{4}\langle c_\perp{}^4\rangle + nm\langle c_\perp{}^2\rangle\,\mathbf{b}\times\frac{D\mathbf{V}}{Dt} + \mathbf{b}\times\frac{\mathbf{n}}{R}nm[\langle c_\parallel{}^2c_\perp{}^2\rangle - \langle\tfrac{1}{4}c_\perp{}^4\rangle]\right\}. \qquad \ldots\ldots(4.59)$$

$D\mathbf{V}/Dt$ may be eliminated through the equations of motion and the flux of $\frac{1}{2}mc_\perp{}^2$ and $\frac{1}{2}mc_\parallel{}^2$ for each component becomes

$$\mathbf{q}_\perp = \frac{1}{eB}\left\{\mathbf{b}\times\nabla\frac{nm^2}{4}\langle c_\perp{}^4\rangle - \frac{m^2}{m_T}\langle c_\perp{}^2\rangle\,\mathbf{b}\times\left[\nabla p_\perp + \frac{\mathbf{n}}{R}p_\parallel\right] + \frac{\mathbf{b}\times\mathbf{n}}{R}nm^2[\langle c_\parallel{}^2c_\perp{}^2\rangle - \tfrac{1}{4}\langle c_\perp{}^4\rangle]\right\} - \frac{\mathbf{j}}{e}\frac{m^2}{m_T}\langle c_\perp{}^2\rangle, \qquad \ldots\ldots(4.60)$$

$$\mathbf{q}_\parallel = \frac{1}{eB}\left\{\mathbf{b}\times\nabla\frac{nm^2}{2}\langle c_\parallel{}^2c_\perp{}^2\rangle - \frac{m^2}{m_T}\langle c_\parallel{}^2\rangle\,\mathbf{b}\times\left[\nabla p_\parallel + \frac{\mathbf{n}}{R}p_\parallel\right] + \frac{\mathbf{b}\times\mathbf{n}}{R}nm^2[\langle c_\parallel{}^4\rangle - \tfrac{3}{2}\langle c_\parallel{}^2c_\perp{}^2\rangle]\right\} - \frac{\mathbf{j}}{e}\frac{m^2}{m_T}\langle c_\parallel{}^2\rangle, \qquad \ldots\ldots(4.61)$$

where $m_T = m_+ + m_-$, while m, e and quantities such as $\langle c_\perp{}^4\rangle$ must be evaluated for the relevant component. The heat fluxes involve higher moments of the distribution function f_0, and no simple relation is available. If, however, f_0 is Gaussian, i.e.

$$f_0 \sim \exp\left\{-\left[\frac{mc_\perp{}^2}{2kT_\perp} + \frac{mc_\parallel{}^2}{2kT_\parallel}\right]\right\},$$

$$\mathbf{q}_-{}^\perp = -\frac{2nkT_\perp{}^-}{eB}\mathbf{b}\times\nabla kT_\perp{}^- - \frac{2kT_\perp{}^-}{eB}\mathbf{b}\times\left[\nabla p_\perp{}^- + \frac{\mathbf{n}}{R}(p_\parallel{}^- - p_\perp{}^-)\right], \qquad \ldots\ldots(4.62)$$

$$\mathbf{q}_+{}^\perp = \frac{2nkT_\perp{}^+}{eB}\mathbf{b}\times\nabla kT_\perp{}^+ - \frac{2kT_\perp{}^+}{eB}\mathbf{b}\times\left[\nabla p_\perp{}^- + \frac{\mathbf{n}}{R}nk(T_\parallel{}^+ - T_\parallel{}^- - T_\perp{}^+)\right] - \mathbf{j}\frac{kT}{e}. \qquad \ldots\ldots(4.63)$$

If the distributions are identical functions of energy, (4.59) clearly implies

$$\mathbf{q}_{\parallel,\perp} = \frac{p_{\parallel,\perp}{}^+}{\Omega_+}\mathbf{b}\times\frac{D\mathbf{V}}{Dt}, \qquad \ldots\ldots(4.64)$$

and in equilibrium $\mathbf{V} = 0$ and $\mathbf{q}$ vanishes to this order.

4.9. *Solutions to the Zero-order Constraint; the Guiding Centre Distribution*

We wish now to determine the properties of the solutions to (4.49). Consider first the steady solutions for which $\mathbf{V} = 0$, $\partial/\partial t = 0$ and (4.49) reduces to

$$c_\parallel\frac{\partial f_0}{\partial x_\parallel} + \frac{e}{m}E_\parallel\frac{\partial f_0}{\partial c_\parallel} - c_\parallel c_\perp{}^2\frac{1}{B}\frac{dB}{dx_\parallel}\left(2\frac{\partial f_0}{\partial c_\parallel{}^2} - \frac{\partial f_0}{\partial c_\perp{}^2}\right) = 0. \qquad \ldots\ldots(4.65)$$

We can find the solution to this by recalling that for this case there exist two constants of the motion, the magnetic moment μ, $\mu = \frac{1}{2}mc_\perp^2/B$ and the energy

$$\mathscr{E} = \tfrac{1}{2}m(c_\parallel^2 + c_\perp^2) - e\int E_\parallel \, dx_\parallel . \qquad \ldots\ldots(4.66)$$

If $E_\parallel$ is derivable from a potential, as is usually the case, the latter becomes $\mathscr{E} = \frac{1}{2}m(c_\parallel^2 + c_\perp^2) + e\phi$. Following § 4.4 we expect that an arbitrary function of these two constants of motion might satisfy (4.65), i.e. $f = f_0(\mu, \mathscr{E}, x_\perp)$. On substituting we obtain

$$\begin{aligned}\left[c_\parallel \frac{\partial \mu}{\partial x_\parallel} + c_\parallel \frac{\mu}{B}\frac{\partial B}{\partial x_\parallel}\right]_c \frac{\partial f_0}{\partial \mu}\bigg|_{\mathscr{E}} + \left[c_\parallel \frac{\partial \mathscr{E}}{\partial x_\parallel} + c_\parallel eE_\parallel - c_\parallel \mu \frac{\partial B}{\partial x}\right]_c \frac{\partial f_0}{\partial \mathscr{E}}\bigg|_\mu \\ = \left[-c_\parallel \frac{\mu}{B}\frac{\partial B}{\partial x_\parallel} + c_\parallel \frac{\mu}{B}\frac{\partial B}{\partial x_\parallel}\right]\frac{\partial f_0}{\partial \mu}\bigg|_{\mathscr{E}} + \left[-c_\parallel eE_\parallel + c_\parallel \mu \frac{\partial B}{\partial x} + c_\parallel eE_\parallel - c_\parallel \mu \frac{\partial B}{\partial x}\right]\frac{\partial f_0}{\partial \mathscr{E}} \\ = 0, \qquad \ldots\ldots(4.67)\end{aligned}$$

where we have used $\mathscr{E} = \frac{1}{2}mc_\parallel^2 + B\mu + e\phi$. This solution, $f_0(\mu, \mathscr{E}, x_\perp)$ is one that we shall find extremely useful.

If the system is not stationary but contains a velocity $\mathbf{V}$ such that $\mathbf{E} + \mathbf{V} \times \mathbf{B} = 0$, we will again try substituting $f_0[\mu(\mathbf{x}, t), \mathscr{E}(\mathbf{x}, t), \mathbf{x}_\perp]$ whereupon (4.49) becomes

$$\begin{aligned}\frac{\partial f_0}{\partial t} + \mathbf{V}.\frac{\partial f_0}{\partial \mathbf{x}}\bigg|_{\mathscr{E},\mu} + \left(\frac{\partial \mathscr{E}}{\partial t} + \mathbf{V}.\frac{\partial \mathscr{E}}{\partial \mathbf{x}}\right)\frac{\partial f_0}{\partial \mathscr{E}}\bigg|_\mu + (\operatorname{div} V)_\parallel \left[(\mathscr{E} - \mu B - e\phi)\frac{\partial f_0}{\partial \mathscr{E}} + \mu \frac{\partial f_0}{\partial \mu}\right] \\ + \left[\frac{\partial \mu}{\partial t} + \mathbf{V}.\frac{\partial \mu}{\partial \mathbf{x}} - \mu \operatorname{div} V\right]\frac{\partial f_0}{\partial \mu} = 0.\end{aligned}$$

The last term here is

$$-\frac{\mu}{B}\left[\frac{\partial B}{\partial t} + \mathbf{V}.\frac{\partial B}{\partial \mathbf{x}} + B \operatorname{div} V\right].$$

This and the third term are negligible if the torsion of the field lines $\partial^2 \mathbf{b}/\partial x_\parallel^2$ is small, whereupon the equation reduces to

$$\frac{Df_0}{Dt}[\mathscr{E}(x, t), \mu]\bigg|_\mu = 0 ; \qquad \ldots\ldots(4.68)$$

i.e. f_0 is a constant on a moving field line.

This first solution $f_0(\mu, \mathscr{E})$, a constant of motion, may be considered in a separate and enlightening fashion. f_0 may be considered in itself as a distribution function for the guiding centres of the orbits, and by itself considered as a first-order specification of the distribution function. It is, however, specified in a frame moving with the guiding centre drift ; hence in calculating moments two precautions must be used. If we are interested in the moment at a point $\mathbf{r}$, we must recall that $\mathbf{r} = \mathbf{R}_0 + \boldsymbol{\rho}$ where $\boldsymbol{\rho} = (1/\Omega)\mathbf{b} \times \mathbf{c}_\perp$, hence we need the distribution function evaluated not at $\mathbf{r}$, but at $[\mathbf{r} - (1/\Omega)\mathbf{b} \times \mathbf{c}_\perp]$. Next the moments of velocity components involve not the relative velocity $\mathbf{c}_\perp$ but the total velocity $\mathbf{V}_0 + \mathbf{c}$. The conditions used here are less restrictive than might appear, and the simplest form of solution is adequate for the description of many plasma equilibria. Instead of exploring this branch of the subject, however, we will turn to the consideration of a useful approximation procedure for dealing with dynamic perturbations.

4.10. *Perturbation Theory*

We shall now attempt to develop a self-consistent perturbation treatment of the Vlasov equation to deal with small disturbances about an equilibrium configuration which is assumed known.

Equilibrium can be obtained by solving the magnetohydrodynamic equations (4.54). The transverse and longitudinal electric fields remain to be determined. $\mathbf{E}_\perp$ can be found by insisting that the plasma be at rest. Thus $\mathbf{E}$ is determined by

$$\rho_+[\mathbf{V}_E + \mathbf{v}_+^{(1)}] + \rho_-[\mathbf{V}_E + \mathbf{v}_-^{(1)}] = 0$$

where $\mathbf{v}^{(1)}$ is the velocity calculated from (4.53) with $D\mathbf{V}/Dt = 0$ a linearizing approximation which assumes all the velocities small. For example, for a cylindrical equilibrium $(1/B)(\partial B/\partial x_\parallel) = 0$, $\mathbf{F} = 0$ and

$$eE_r = \frac{2p_\parallel^+}{n}\frac{B_\theta}{rB_z} + \frac{p_\perp^+}{n}\frac{\partial}{\partial r}(\log B) - \frac{1}{n}\frac{\partial}{\partial r}p_\perp^+. \qquad \ldots\ldots(4.69)$$

When this has been done the problem is reduced to that of solving the magnetohydrodynamic equations

$$\mathbf{j} \times \mathbf{B} = \nabla p + \frac{\mathbf{n}}{R}p_\parallel$$

or

$$(\mathbf{j} \times \mathbf{B})_r = -\frac{\partial p_\perp}{\partial r} - \frac{B_\theta}{rB_z}(p_\parallel - p_\perp). \qquad \ldots\ldots(4.70)$$

If $p_\parallel = p_\perp$ this reduces to the magnetohydrodynamic solution.

Having obtained an equilibrium configuration we must discover how it responds to a small perturbation about equilibrium. Since we are interested in a first approximation to the motion, it is enough to calculate to zero order in r_L. We can therefore make use of Eqn (4.49) and consider a perturbation introduced by a perturbing drift velocity $\mathbf{V}'$. We introduce the distribution function $f = f^0 + f'$ where f' is the perturbation introduced by $\mathbf{V}'$, and the linearized form of (4.49) becomes

$$\left[\frac{Df'}{Dt} + c_\parallel\frac{\partial f'}{\partial x_\parallel} + \frac{e}{m}E_\parallel\frac{\partial f'}{\partial c_\parallel} - c_\parallel\frac{c_\perp^2}{B}\frac{\partial B}{\partial x_\parallel}\left(\frac{\partial f_0'}{\partial c_\parallel^2} - \frac{\partial f_0'}{\partial c_\perp^2}\right)\right]$$

$$-(\operatorname{div} V)_\perp c_\perp^2\frac{\partial f^0}{\partial c_\perp^2} - c_\parallel^2(\operatorname{div} V)_\parallel\frac{\partial f^0}{\partial c_\parallel^2} = 0,$$

or, on introducing the variables μ and $\mathscr{E}$,

$$\mathbf{O}f' = +(\operatorname{div} V)_\perp\mu\frac{\partial f_0}{\partial \mu} + (\operatorname{div} V)_\parallel\left[\mu\frac{\partial f_0}{\partial \mu} - (\mathscr{E} - B_\mu - c\phi)\frac{\partial f}{\partial \mathscr{E}}\right]. \qquad \ldots\ldots(4.71)$$

Now, the operator on the left describes the motion of a particle moving along the field line with a velocity $c_\parallel(\mathscr{E}, \mu)$; hence the solution of (4.71) may be written

$$f' = \int dt' \operatorname{div} V\mu\frac{\partial f_0}{\partial \mu} + (\operatorname{div} V)_\parallel\left[\mu\frac{\partial f_0}{\partial \mu} - (\mathscr{E} - B\mu - e\phi)\frac{\partial f_0}{\partial \mathscr{E}}\right]. \qquad \ldots\ldots(4.72)$$

This, however, represents the *convected* perturbation and to return it to the original location we must subtract $\boldsymbol{\xi}.\nabla f_0$ where $\partial\boldsymbol{\xi}/\partial t = \mathbf{V}$. In terms of the variable $\boldsymbol{\xi}$

$$f' = -(\boldsymbol{\xi}.\nabla)f_0 + \int dt' \left\{ \operatorname{div}\dot{\boldsymbol{\xi}}\mu \frac{\partial f_0}{\partial \mu} + (\operatorname{div}\boldsymbol{\xi})_{\parallel}\left[\mu\frac{\partial f_0}{\partial \mu} - 2(\mathscr{E} - B\mu - e\phi)\frac{\partial f_0}{\partial \mathscr{E}}\right]\right\}. \qquad \ldots\ldots(4.73)$$

The integral along the trajectory may be written

$$\int dt' = \int \frac{dx_{\parallel}}{c_{\parallel}} = \sqrt{\left(\frac{2}{m}\right)} \int \frac{dx_{\parallel}}{\sqrt{(\mathscr{E} - B\mu - e\phi)}} \qquad \ldots\ldots(4.74)$$

and here a fundamental difficulty presents itself. Generally the potential ϕ will be altered during a perturbation, and while it is tempting to expand in powers of the disturbance in ϕ, this introduces serious difficulties near any turning points in the orbits, since an attempt to expand (4.73) will introduce a divergence. However, the quantity ϕ may be determined by insisting on quasi-neutrality. In evaluating the integrals over μ and ϵ which are needed to form n, we may integrate by parts and, at the expense of needing ever higher derivatives of the distribution function, obtain an expression relating $\delta\phi$ to the $\boldsymbol{\xi}$'s. Finally, an equation of motion is needed and here we use the magnetohydrodynamic equations (4.51–53), but now we must calculate the pressure tensor from (4.73). Since, in our approximation, the plasma behaves like a perfect conductor, we may use

$$0 = \mathbf{E} + \dot{\boldsymbol{\xi}} \times \mathbf{B}_0 \quad \text{and} \quad \frac{\partial \mathbf{B}'}{\partial t} = -\operatorname{curl}\mathbf{E}\,; \quad \mathbf{B}' = \operatorname{curl}(\boldsymbol{\xi} \times \mathbf{B}_0) = \mathbf{Q}$$

to obtain the fields, then write the equation of motion in a form resembling (1.15), differing from it only in that the pressure tensor must be calculated from (4.73).

$$\rho_0 \ddot{\boldsymbol{\xi}} = -\nabla.\mathbf{p}' + \mathbf{j} \times \mathbf{Q} + \operatorname{curl}\mathbf{Q} \times \mathbf{B}_0, \qquad \ldots\ldots(4.75)$$

where

$$p_{ij}' = \int f' c_i c_j.$$

There are two distinct methods of handling the time integrals in (4.73). For finite equilibrium systems the time integrals are periodic, either because particles are trapped in some region of the field or because orbits following field lines close on themselves. If we attempt to solve a stability problem, we shall consider displacements $\boldsymbol{\xi}$ whose time dependence is represented by a Laplace transform

$$\boldsymbol{\xi}(t) = \int \boldsymbol{\xi}(s)\, e^{st}\, ds$$

where $R(s) > 0$, and if we are interested in marginal stability we will seek solutions for which s is small. We can then handle the integrals by observing that for a periodic function (Rosenbluth and Rostoker 1959)

$$\int f e^{st} \simeq \frac{1}{1 - e^{-s\tau}} \int_{t-\tau}^{t} f e^{st}\,;$$

hence the integrals become

$$\frac{s}{1-e^{-s\tau}}\oint f$$

$$\simeq \frac{1}{\tau}\oint f \quad \text{if} \quad s \ll \tau\,;$$

and the pressure involves not the time derivative of the displacement $\boldsymbol{\xi}$, but the displacement $\boldsymbol{\xi}$ itself.

The period τ here is

$$\oint \frac{dx_\parallel}{\sqrt{(\mathscr{E}-B\mu-e\phi)}},$$

where the integral is closed at the zeros of the square root if particles are trapped, or with the period of the system if flux lines are closed.

The problem can be simplified still further by assuming that: (*a*) $e\phi$ is negligible; (*b*) the distribution function f_0 is isotropic.

If we write $c_\perp{}^2 = c^2 \sin^2\theta$ at the point x, then in the integrals introduce

$$c_\perp{}^2(x') = c_\perp{}^2(x)\frac{B(x')}{B(x)} = c^2\sin^2\theta\frac{B(x')}{B(x)}$$

to give

$$\frac{1}{\tau}\oint \frac{dx_\parallel}{c_\parallel\sqrt{\{1-B(x')\sin^2\theta/B(x)\}}}\operatorname{div}\xi + (\operatorname{div}\xi)_\parallel\, 2\left[1-\tfrac{3}{2}\sin^2\theta\frac{B(x')}{B(x)}\right]c^2\frac{\partial f_0}{\partial c^2} = A(\xi)c^2\frac{\partial f_0}{\partial c^2}. \qquad \ldots\ldots(4.76)$$

But since

$$\tau = \oint \frac{dx_\parallel}{c\sqrt{[1-B(x')\sin^2\theta/B(x)]}}$$

the velocity variables are removed from under the integrals in A, which has a purely geometric character. If, for example, the unperturbed field does not vary in magnitude along a field line, these expressions become

$$\frac{A}{\tau}c^2\frac{\partial f_0}{\partial c^2} = \{\langle\operatorname{div}\boldsymbol{\xi}\rangle + \tfrac{1}{2}\langle(\operatorname{div}\boldsymbol{\xi})_\parallel\rangle\}\,c^2\frac{\partial f_0}{\partial c^2},$$

where

$$\langle\operatorname{div}\boldsymbol{\xi}\rangle = (1/L)\oint dx_\parallel \operatorname{div}\boldsymbol{\xi}$$

is the average of the divergence of the displacement taken around a field line.

To form the pressure tensor we must multiply by the square of the components $c_\perp{}^2$, or $c_\parallel{}^2$ and since $c_\perp{}^2 = c^2\sin^2\theta$, $c_\parallel{}^2 = c^2\cos^2\theta$, the perturbed distribution function is anisotropic, and the pressure tensor becomes

$$\mathbf{p}' = \boldsymbol{\xi}.\nabla p_0\,\mathbf{1} + p_\perp{}'(\mathbf{1}-\mathbf{bb}) + p_\parallel{}'\,\mathbf{bb},$$

where $p_\perp{}'$ and $p_\parallel{}'$ are

$$p_\perp{}' = p_0\int\frac{A(\theta)}{\tau}\sin^2\theta\,d\theta,$$

$$p_\parallel{}' = p_0\int\frac{A(\theta)}{\tau}\cos^2\theta\,d\theta. \qquad \ldots\ldots(4.77)$$

For a system in which the field strength does not vary along a line of force, these become

$$p_\perp' = p_0[\langle \operatorname{div} \xi\rangle - \tfrac{1}{4}\langle \operatorname{div} \xi\rangle_\parallel]$$
$$p_\parallel' = p_0[\langle \operatorname{div} \xi\rangle + \tfrac{3}{4}\langle \operatorname{div} \xi\rangle_\parallel]. \qquad \ldots\ldots(4.78)$$

It is possible again to form an energy integral from the equations of motion. We obtain for δW

$$\delta W = \delta W_{\mathrm{MH}} - \gamma \int p(\operatorname{div} \xi)^2 - \int \xi . \nabla . [\mathbf{p}' - \xi . \nabla p . \mathbf{1}] \qquad \ldots\ldots(4.79)$$
$$= \delta W_1 + \int (\mathbf{p}' . \nabla) . \xi,$$

where δW_{MH} is given by (1.32), $\delta W_1 = \delta W_{\mathrm{MH}} - \gamma \int p(\operatorname{div} \xi)^2$; and

$$\int (\mathbf{p}' . \nabla) . \xi = \int \{p_\perp' \operatorname{div} \xi + (p_\parallel' - p_\perp')(\mathbf{b} . \nabla)\xi . \mathbf{b}\}$$
$$= \int \{p_\perp' \operatorname{div} \xi + (p_\parallel' - p_\perp')(\operatorname{div} \xi)_\parallel\}.$$

This again is a quadratic functional of the ξ's, and it may be shown positive definite (Rosenbluth and Rostoker 1959), while the equations of motion take the form

$$\int \rho \ddot{\xi}\, d\tau = -\frac{\partial(\delta W)}{\partial \xi},$$

where $\partial(\delta W)/\partial \xi$ is the functional derivative of δW. Hence, this constitutes the energy and it must be minimized if the configuration is to be in equilibrium.

Rosenbluth and Rostoker (1959) have shown that the energy defined above is greater than that calculated from the magnetohydrodynamic equations, and less than that calculated from the double-adiabatic equations, where $\mathbf{p}$ is given by (4.57); hence if a system is magnetohydrodynamically stable, it is stable in this theory, while if unstable in the double-adiabatic approximation it is unstable in this theory (see also Kruskal and Oberman 1958).

This reasonably sophisticated theory of magnetohydrodynamic stability is however not quite complete, in two senses. In the first place, the role of the longitudinal electric field has been underplayed, and if its effects are properly considered the problem becomes much more complex. The quasi-neutrality condition, $n_+ = n_-$, must first be used to determine this field, which then must be used to determine the perturbed pressures. Newcomb has given a careful study of this, which indicates that two separate quantities, one associated with the perturbation and one with the longitudinal electric field, must be minimized.

A possibly more serious omission has been the neglect of all small scale instabilities. In the treatment given here we have used an expansion in powers of $\epsilon = (r_L/B)(\partial B/\partial x)$, for both the equilibrium condition and the perturbation, i.e. not only is ϵ small for the equilibrium, but for a perturbation $\xi(k, t)$, $r_L k$ is small. There may, however, be terms in the distribution function which are not analytic in this quantity, e.g. terms such as $\exp(-1/\epsilon)$, and which could not contribute to our expansion. Also, for small scale disturbances for which $\partial\xi/\partial x \sim \xi/r_L$, the

expansion cannot be expected to converge. Hence in the quasi-hydrodynamic theory presented here small scale perturbations are not properly treated.

By analogy with the uniform two-stream problem, one might expect difficulties to arise from the Landau damping terms, and indeed, near cyclotron resonances, these terms have such forms as

$$\exp\left\{-\left(\frac{\omega}{kv_\theta}\right)^2\right\} \simeq \exp\left\{-\left(\frac{\omega}{\Omega k r_L}\right)^2\right\}$$

which is certainly not analytic in kr_L. Some attempts have been made to discuss these instabilities, using the method of § 4.1, but these prove even more complex than the quasi-hydrodynamic analysis. However, it is found that there are instabilities associated with the Landau terms. In a magnetic field there are two such terms, one associated with motion along field lines, coming from integrals of the form

$$\int \frac{f(v_\parallel)\,dv_\parallel}{(\omega + n\Omega + k_\parallel v_\parallel)}$$

and others arising only if the wave vector forms a finite angle with the field and arising from terms such as

$$\int \frac{f(v_\parallel)\,dv_\parallel}{(\omega + k_\parallel v_\parallel)^2 - \Omega^2}.$$

The first of these usually corresponds to some version of normal two-stream instability, but the second involves a resonance between the plasma frequency and a gyrofrequency. A simple example was discussed by Harris (1959) who considered a spatially homogeneous distribution with an anisotropic velocity distribution and showed that for a great enough anisotropy the distribution would become unstable provided that the electron plasma frequency was greater than the ion gyrofrequency, i.e. $1 > (m_-/m_+)(B^2/4\pi\rho c^2)$. He found the growth rate to be approximately $[(\Omega_+/120) - k_\parallel v_\parallel{}^\theta]$ where $(v_\parallel{}^\theta)^2$ is the mean square value of the parallel component of velocity and is assumed much smaller than $(v_\perp{}^\theta)^2$.

Similar instabilities can be discovered in the double adiabatic approximation where $p_\parallel - p_\perp \neq 0$, and more may be expected.

4.11. *Theory of Sound*

It is instructive to consider the propagation of low frequency oscillations through a uniform plasma in a uniform field. In this case, the disturbance may be Fourier analysed in space and time as $\boldsymbol{\xi} = \Sigma\boldsymbol{\xi}(\mathbf{k}, \omega)\exp[i(\mathbf{k}.\mathbf{x} + \omega t)$, the trajectory integrals become $1/i(k_\parallel c_\parallel + \omega)$, with the Landau prescription, and the perturbed distribution function becomes

$$f' = \frac{i\omega}{i(\omega + k_\parallel c_\parallel)}\left\{(\operatorname{div}\boldsymbol{\xi})\,\mu\frac{\partial f_0}{\partial \mu} + (\operatorname{div}\boldsymbol{\xi})_\parallel\left[\mu\frac{\partial f_0}{\partial \mu} - (\mathscr{E} - 2B\mu)\frac{\partial f_0}{\partial \mathscr{E}}\right]\right\}$$

or, in terms of the velocities,

$$f' = \frac{\omega}{\omega + k_\parallel c_\parallel}\left[(\operatorname{div}\boldsymbol{\xi})_\perp c_\perp{}^2\frac{\partial f_0}{\partial c_\perp{}^2} + (\operatorname{div}\boldsymbol{\xi})_\parallel c_\parallel\frac{\partial f_0}{\partial c_\parallel}\right]. \quad \ldots\ldots(4.80)$$

The components of the pressure tensor,

$$p_{\parallel} = \int f' m c_{\parallel}^2 \, d^3c$$

$$p_{\perp} = \int f' \tfrac{1}{2} m c_{\perp}^2 \, d^3c, \qquad \ldots\ldots(4.81)$$

may then be written down, and for high phase velocity take on a particularly simple form ; for if $\omega \gg k_{\parallel} c_{\parallel}$ then

$$\int \frac{1}{\omega + k_{\parallel} c_{\parallel}} g(c) \simeq \frac{1}{\omega} \int g(c),$$

and

$$p_{\parallel}' = p_{\parallel}^0 [(\operatorname{div} \boldsymbol{\xi})_{\perp} + 3(\operatorname{div} \boldsymbol{\xi})_{\parallel}]$$

$$p_{\perp}' = -p_{\perp}^0 [2(\operatorname{div} \boldsymbol{\xi})_{\perp} + (\operatorname{div} \boldsymbol{\xi})_{\parallel}]$$

which are the double adiabatic relations (cf. (4.57)). Under these circumstances, the equation of motion for $\boldsymbol{\xi}$ has the form

$$-\omega^2 \rho_0 \boldsymbol{\xi} = -\nabla p' + (\mathbf{j}' \times \mathbf{B}_0) \qquad \ldots\ldots(4.82)$$

where

$$\mathbf{j}' = \frac{1}{4\pi} \operatorname{curl} \mathbf{B}',$$

$$\mathbf{B}' = -\mathbf{B}_0 (\operatorname{div} \boldsymbol{\xi}) + (\mathbf{B}_0 . \nabla) \boldsymbol{\xi}$$

and

$$p_{ij}' = p_{\perp}' \delta_{ij} + (p_{\parallel}' - p_{\perp}') b_i^0 b_j^0 + (p_{\parallel}^0 - p_{\perp}^0) [\delta b_i b_j^0 + b_i^0 \delta b_j],$$

$\delta \mathbf{b}$ being the perturbation in the unit vector along the perturbed magnetic field,

$$\delta \mathbf{b} = (\mathbf{b}_0 . \nabla) \boldsymbol{\xi}_{\perp}.$$

The components of (4.82) become

$$(\rho_0 \omega^2 - 3 p_{\parallel} k_{\parallel}^2) \xi_{\parallel} = p_{\perp} k_{\parallel} (\mathbf{k}_{\perp} . \boldsymbol{\xi}_{\perp}) \qquad \ldots\ldots(4.83)$$

$$\left[\rho_0 \omega^2 + k_{\parallel}^2 \left(p_{\parallel} - p_{\perp} - \frac{B^2}{4\pi}\right)\right] \boldsymbol{\xi}_{\perp} = \mathbf{k}_{\perp} \left[2(\mathbf{k}_{\perp} . \boldsymbol{\xi}_{\perp}) \left(p_{\perp} + \frac{B^2}{8\pi}\right) + k_{\parallel} p_{\perp} \xi_{\parallel}\right].$$

We may now introduce the Alfvén speed, $V_A^2 = B^2/4\pi\rho$; the parallel and perpendicular ' sound speeds ', $V_{\parallel}^2 = p_{\parallel}/\rho$, $V_{\perp}^2 = p_{\perp}/\rho$; the phase velocity, $V = -\omega/k$; and the direction of propagation θ such that $k \cos\theta = k_{\parallel}$. The displacement $\boldsymbol{\xi}$ may be eliminated from (4.83) leaving the dispersion relation

$$(V^2 - 3V_{\parallel}^2 \cos^2\theta)[V^2 - (V_A^2 - V_{\perp}^2 - V_{\parallel}^2)\cos^2\theta - (2V_{\perp}^2 + V_A^2)\sin^2\theta] - V_{\perp}^4 \sin^2\theta \cos^2\theta = 0. \qquad \ldots\ldots(4.84)$$

This may be factorized if $\theta = 0$ or $\pi/2$. For parallel propagation, $\theta = 0$,

$$V^2 = 3V_{\parallel}^2$$

or

$$V^2 = V_A^2 + V_{\perp}^2 - V_{\parallel}^2. \qquad \ldots\ldots(4.85)$$

If

$$V_{\parallel}^2 > V_A^2 + V_{\perp}^2$$

i.e.

$$p_{\parallel} > p_{\perp} + \frac{B^2}{4\pi}$$

the second root becomes imaginary, and instead of oscillations, instabilities occur. These are the ' firehose ' instabilities and arise from the centrifugal accelerations experienced by particles streaming along curved but flexible paths.

For perpendicular propagation $\theta = \pi/2$

$$V^2 = 2V_\perp^2 + V_A^2, \quad \text{or} \quad V^2 = 0, \qquad \ldots\ldots(4.86)$$

which correspond to the fast and slow magnetosonic waves.

At arbitrary angles, the full quadratic must be solved. Writing (4.84) as $V^4 + bV^2 + c = 0$, the roots are

$$V^2 = \frac{-b \pm (b^2 - 4c)^{1/2}}{2}, \quad \text{with} \quad b < 0.$$

A negative root will appear if $c < 0$, hence there will be instabilities if

$$V_\perp^4 \sin^2\theta > 3V_\parallel^2[V_A^2 + V_\perp^2(1 + \sin^2\theta) - V_\parallel^2 \cos^2\theta]. \qquad \ldots\ldots(4.87)$$

For propagation almost at right angles to the field this condition becomes

$$V_\perp^4 > 3V_\parallel^2[V_A^2 + 2V_\perp^2]$$

or

$$\frac{1}{6}\beta_\perp \frac{p_\perp}{p_\parallel} > \beta_\perp + \frac{B^2}{8\pi}. \qquad \ldots\ldots(4.88)$$

This is the ' mirror ' instability and arises from the trapping of particles in regions of weak field. If the field is locally weak, particles are retarded by the acceleration $\mu(\partial B/\partial x)$, the local pressure increases and the magnetic field is still further weakened.

A more detailed investigation, including the effect of the longitudinal field $E_\parallel$, and a more careful treatment of the trajectory integrals alters this picture quantitatively, and qualitatively through the appearance of Landau damping, which attenuates waves other than the transverse Alfvén wave, propagating along the magnetic field lines ; however, except for that modification, the qualitative aspects of this section are unaltered, and the gross features of the double adiabatic system remain.

§ 5. Conclusion

After this somewhat formal excursion it is worth while looking back to see what has been achieved.

The magnetohydrodynamic theory of the ideal plasma has been rather fully developed, but its applicability depends on the flow occurring at large magnetic Reynolds numbers, i.e.

$$\frac{\mu\sigma}{c^2} LB \sqrt{\left(\frac{4\pi}{\rho}\right)} \gg 1.$$

In astrophysical systems this can usually be satisfied since the distance scales L are enormous ; however, on the laboratory scale it is well satisfied only for hot rarefied plasmas. In these, the mean free path increases with temperature as T^2, hence in hot plasmas the mean free path is long. Within the context of classical kinetic theory, a long mean free path means an imperfect fluid, since transport coefficients are of order λ, and our first object was to examine the real fluid equations produced by the classical kinetic theory.

In the course of doing this two limitations were observed. In the first place, the entire basis of the theory, Boltzmann's equation, was derived for a diffuse system interacting through short range forces while, for the Coulomb interaction, the forces are long range and the total effective collision cross section

$$Q \sim \int \frac{(1-\cos^2\theta)}{(\sin^4\theta/2)}\, d\Omega$$

diverged, as $\log(\sin\frac{1}{2}\theta)$. Secondly, the solution procedure, based as it was on an expansion valid for small collision intervals, was inapplicable to a diffuse plasma, where magnetic field variations can introduce rapid fluctuations and the long mean free path keeps down the collision frequencies. In particular, the gyrofrequency Ω of particles may be very much greater than the collision frequency $1/\tau$, and while some of the transport coefficients went to reasonable limits as $\Omega\tau \to \infty$, this was not true of all.

We therefore turned to a treatment of plasma dynamics in which collisions were omitted altogether, in the belief that such a treatment would be useful in studying processes which occurred in times short compared with a collision period, and might act as a basis upon which a more realistic treatment of collisions could be based. The first result here was a more sophisticated calculation of the dielectric coefficient at short wavelengths, a calculation which has little bearing on the actual propagation of long wavelength oscillations through the plasma, but which did enable us to give a fairly rational treatment of the effective interaction in a plasma. For a steady homogeneous plasma, we reproduced the dominant terms in the screened Boltzmann equation, but in non-uniform plasmas the demonstration is incomplete since we do not fully understand the effect of inhomogeneities on interaction by means of travelling plasma waves, although it is usually small except for systems which are unstable against growing plasma oscillations.

We then turned to consider the much more difficult problem of the behaviour of a collisionless plasma in a magnetic field. Here we managed to produce a form for the dielectric coefficient which had a sensible behaviour in the long wavelength region, although the form was elaborate rather than intelligible.

Finally we reached the problem which was originally posed : that of finding the collisionless analogue to magnetohydrodynamics. Using an expansion which required Ω to be much greater than any hydrodynamic frequency, we were able to produce an approximate solution to the Vlasov equation, which in the limit reduced to an equation describing the motion along field lines of electrically charged magnetic dipoles, and became a one-dimensional equation. For this, a perturbation treatment was developed, and the analogues of sound waves were produced. There were three of these, the Alfvén wave, and a fast and slow sound wave which had many of the features of magnetohydrodynamic sound. A characteristic of this motion is that the pressure is determined not by local values of the variables but by mean values along lines of force. A rather simplified stability analysis could be carried out and again the process has a considerable resemblance to magneto-hydrodynamics, particularly for interchange perturbations where there is little variation along a field line.

We have given some slight attention to the non-ideal aspects of the collisionless plasma. The earliest imperfections to arise occur in the pressure tensor and in the

current. Even in the collisionless limit there is a ' viscosity ', although since its effect is rotated, it leads to no dissipation ; none the less, it may have some effect on the finer details of stability calculations—besides raising the order of the differential equations. In the collisionless gas, the flux is convected with the velocity $\mathbf{V}_E$; on the other hand, matter is convected with a velocity $\mathbf{V}_D = \mathbf{V}_E + \mathbf{V}_N$ where the second term arises because of curvature in the field lines and depends on the sign of the charge; thus even although there are no collisions, the fluid does not behave quite like an ideal conductor, since there is a slip of order $1/\Omega T$ between the fluid and the magnetic flux.

What bearing does the theory developed in here have on reality? To a large extent the connections are yet to be displayed. There is adequate evidence for the grosser aspects of magnetohydrodynamics ; magnetic flux has been shown to be swept along with the fluid, even when the fluid is accelerated by non-electric means. There is also little doubt that a magnetic field can compress a fluid, and that rather good agreement can be achieved between calculated and observed rates and degree of collapse. The theory of sound is less satisfactory, although Alfvén waves have been demonstrated in a plasma. Some of the grosser predictions of stability theory have been checked, and it appears that plasmas predicted by theory to be unstable are in fact unstable.

Some experiments have been performed in an attempt to produce stable configurations, but with less success. Instabilities are detected as random fluctuations on the magnetic fields measured by probing coils, and while these are less intense and longer in appearing than in theoretically unstable configurations they do appear. This rather resembles the stability of hydrodynamic flow, in which experimentally it is difficult to realize the theoretical stability limits.

A further theoretical complication lies in the occurrence of an increasing number of new types of instabilities, one associated it seems with each refinement in the theory ; and this points in the direction where the analysis presented in this review may break down. It appears as a logical method to approach the study of the diffuse plasma, but conceivably the ideal steady, stable and intelligible system with which the theory can cope may never arise in nature because of all-prevailing instabilities, which make all plasmas turbulent. If this is the case, and there is little evidence against it, the plasma is even less intelligible than we have thought. We must hope, however, that a stable hot diffuse plasma, to which the theory here presented applies, can exist in a magnetic field, and look forward to its production in the laboratory.

Acknowledgments

It is a pleasure to acknowledge the usefulness of discussions with my colleagues at Harwell, and with our consultants. Special thanks go to R. J. Tayler and E. W. Laing who corrected many errors in the manuscript.

References

The following references have been made: *The Proceedings of the 2nd International Conference on the Peaceful Uses of Atomic Energy, Geneva*, is abbreviated to *Geneva*, and *The Proceedings of the Fourth International Conference on Ionization Phenomena in Gases, Uppsala*, is abbreviated to *Uppsala*.

AITKEN, K. L., BURCHAM, J. M., and REYNOLDS, P., 1960, *Uppsala* (Amsterdam : North-Holland), p. 896.
ALFVÉN, H., 1950, *Cosmical Electrodynamics* (Oxford : University Press).
ARTSIMOVICH, L., 1958, *Geneva*, **31**, 6 ; 1959, *Progr. Nucl. Energy*, [XIII], **1**, 6.
AUER, P. L., 1958, *Phys. Rev. Letters*, **1**, 411.
BABCOCK, H. W., 1947, *Astrophys. J.*, **105**, 105.
BENNETT, W. H., 1934, *Phys. Rev.*, **45**, 890.
BERNSTEIN, I. B., 1958, *Phys. Rev.*, **109**, 10.
BERNSTEIN, I. B., FRIEMAN, E. A., KRUSKAL, M. D., and KULSRUD, R. M., 1958, *Proc. Roy. Soc.* A, **244**, 17.
BIRDSALL, D. H., COLGATE, S. A., and FÜRTH, H., 1960, *Uppsala* (Amsterdam : North-Holland), pp. 888, 892.
BOHM, D., and PINES, D., 1952, *Phys. Rev.*, **85**, 338.
BROWN, S. C., 1960, *Uppsala* (Amsterdam : North-Holland), p. 691.
BUNEMAN, O., 1959, *Phys. Rev.*, **115**, 503.
BURKHARDT, L. C., and LOVBERG, R. H., 1958, *Geneva*, **32**, 29; 1959, *Progr. Nucl. Energy*, [XIII], **1**, 405.
CARRUTHERS, R., and DAVENPORT, P. A., 1957, *Proc. Phys. Soc.* B, **70**, 49.
CHANDRASEKHAR, S., KAUFMAN, A. N., and WATSON, K. M., 1957, *Ann. Phys., N.Y.*, **2**, 435 ; 1958, *Ann. Phys., N.Y.*, **5**, 1.
CHAPMAN, S., and COWLING, T. G., 1951, *The Mathematical Theory of Non-uniform Gases*, 2nd Edition (Cambridge : University Press).
CHEW, G., GOLDBERGER, M. L., and LOW, F. E., 1956, *Proc. Roy. Soc.* A, **236**, 12.
COWLING, T. G., 1957, *Magnetohydrodynamics* (New York : Interscience).
CURZON, F. L., FOLKIERSKI, A., LATHAM, R., and NATION, J. A., 1960, *Proc. Roy. Soc.* A, **257**, 386.
FERMI, E., 1949, *Phys. Rev.*, **75**, 1169.
FERRARO, V. C. A., 1937, *Mon. Not. R. Astr. Soc.*, **97**, 458.
FOWLER, R. H., 1936, *Statistical Mechanics*, 2nd Edition (Cambridge : University Press).
FRIEDRICHS, K. O., and KRANZER, H., 1958, *Notes on Magnetohydrodynamics VIII, Non-linear Wave Motion*, New York University Report N.Y.O.6486.
GREEN, H. S., 1952, *The Molecular Theory of Fluids* (Amsterdam : North-Holland).
HALE, G. E., 1908, *Astrophys. J.*, **28**, 28.
HARRIS, E. G., 1957, *Naval Research Laboratory Report* 4944 ; 1959, *Proc. Conf. on Plasma Oscillations* (Indianapolis : Linde Company), p. 69.
HILBERT, D., 1912, *Math. Ann.*, **7**, 562.
HILTNER, W. A., 1950, *Phys. Rev.*, **78**, 150.
HIRSCHFELDER, J. D., CURTIS, C. F., and BIRD, R. B., 1954, *Molecular Theory of Gases and Liquids* (New York : Wiley).
KHARCHENKO, I. F., FAINBERG, YA. B., NIKOLAYEV, P. M., KORNILOV, E. A., LUTZENKO, E. A., and PEDENKO, N. S., 1960, *Uppsala* (Amsterdam : North-Holland), p. 671.
KRUSKAL, M. D., and KULSRUD, R. M., 1958, *Phys. Fluids*, **1**, 265.
KRUSKAL, M. D., and OBERMAN, C. R., 1958, *Phys. Fluids*, **1**, 275.
KRUSKAL, M., and SCHWARZSCHILD, M., 1954, *Proc. Roy. Soc.* A, **223**, 348.
LAING, E. W., ROBERTS, S. J., and WHIPPLE, R. T. P., 1959, *J. Nucl. Energy*, **1**, 49.
LANDAU, L., 1937, *J. Exp. Theor. Phys.*, 7, 203; 1946, *J. Phys., U.S.S.R.*, **10**, 25.
LANDSHOFF, R., 1949, *Phys. Rev.*, **76**, 904.
LEES, D. J., and RUSBRIDGE, M. G., 1960, *Uppsala* (Amsterdam : North-Holland), p. 954.
LIGHTHILL, M. J., 1960, *Phil. Trans.* A, **252**, 397.
LIN, C. C., 1955, *The Theory of Hydrodynamic Stability* (Cambridge : University Press).
LÜST, R., and SCHLÜTER, A., 1957, *Z. Naturf.*, **12a**, 850.
MARSHALL, W., 1957, *Atomic Energy Research Establishment Report* T/R 2247 ; 1958, *Atomic Energy Research Establishment Report* 2352, 2419.
NEWCOMB, W. A., 1959, *University of California Radiation Laboratory Report*, 5447.
PENROSE, O., 1960, *Phys. Fluids*, **3**, 258.
POST, R. F., 1956, *Rev. Mod. Phys.*, **28**, 338.

RATCLIFFE, J. A., 1959, *The Magneto-ionic Theory and its Application to the Ionosphere* (Cambridge : University Press).
ROSENBLUTH, M. N., 1958, *Geneva*, **31**, 85 ; 1959, *Progr. Nucl. Energy*, [XIII], **1**, 463.
ROSENBLUTH, M. N., and ROSTOKER, N., 1959, *Phys. Fluids*, **2**, 23.
SHKLOVSKY, I. S., 1953, *Dokl. Akad. Nauk S.S.S.R.*, **90**, 983.
SPARROW, J., 1960, *Atomic Energy Research Establishment Report* R3301.
SPITZER, L., 1956, *The Physics of Fully Ionized Gases* (New York : Interscience).
SUYDAM, B. R., 1958, *Geneva*, **31**, 157 ; 1959, *Progr. Nucl. Energy*, [XIII], **1**, 431.
SYROVATSKI, C. J., 1957, *Usp. Fis. Nauk.*, **62**, 247.
TAYLER, R. J., 1957, *Proc. Phys. Soc.* B, **70**, 1049.
THOMPSON, W. B., 1957, *Nature, Lond.*, **179**, 886.
THOMPSON, W. B., and HUBBARD, J., 1960, *Rev. Mod. Phys.*, **32**, 714.
THONEMANN, P. C., 1958, *Geneva*, **31**, 34 ; 1959, *Progr. Nucl. Energy*, [XIII], **1**, 50.
THONEMANN, P. C., and COWHIG, W. T., 1951, *Proc. Phys. Soc.* B, **64**, 345.
TUCK, J. L., 1958, *Geneva*, **32**, 3 ; 1959, *Progr. Nucl. Energy*, [XIII], **1**, 50.
VAUGHAN-WILLIAMS, R. W., and HAAS, F. A., 1961, *Phys. Rev. Letters*, **6**, 165.
VLASOV, A., 1938, *J. Exp. Theor. Phys.*, **8**, 291.
WATSON, K. M., 1956, *Phys. Rev.*, **102**, 12.
WHIPPLE, R. T. P., 1960, *Atomic Energy Research Establishment Report* R3325.